Savona (Italy)

May 16, 2020

Dear Editor,

We would like to submit the manuscript *Climate elasticity of evapotranspiration during multi-year droughts shifts the water balance of Mediterranean rain-snow climates* for publication in Hydrology and Earth System Sciences. This is a resubmission of manuscript **hess-2019-377**, which was published in open discussion on August 26, 2019 and was reviewed by two referees. In doing so, we would like to thank the Editor and the Editorial Office for their patience and flexibility with our deadlines.

We have revised and improved the manuscript based on comments from all reviewers and would like to thank all of you for taking the time to review our manuscript. Our revisions followed the expected changes we outlined in our public response to referees. We prioritized (1) providing more details about the spatially distributed evapotranspiration product we used in this paper, (2) improving our Introduction to better frame our research questions, and (3) expanding our discussion to clarify climate elasticity of evapotranspiration, including a schematic as suggested by Referee 1. We also included in the manuscript some discussion on why we did not consider a groundwater model or a hypothetical design drought, again following the arguments we provided in our public response. All minor points were fixed following comments of both referees.

Please find attached our point-by-point replies and the new version of our manuscript for details. We also attached a version of the manuscript with tracked changes.

Francesco Avanzi and coauthors

Reply to Reviewer #1

The manuscript by Avanzi et al. is generally well-written and well-referenced. They utilize a hydrologic modeling approach to quantify how a mountain watershed responds to short duration (sub-decadal) drought episodes, and find that how simulated evapotranspiration responds to drought conditions is a major issue with regards to runoff estimation. While unsurprising, this finding is important to motivate improvements in ET in hydrologic models in mountain regions. Overall, the modelling approach is acceptable, but will need some additional information and analysis (noted below), especially with regards to the ET estimation approach and some additional longer simulations.

I am also a bit skeptical of the fact that a groundwater model was not used to show whether or not such a model is necessary to accurately capture hydrologic responses in volcanic, subsurface flow-dominated basins. I liked the shift identification approach for changes in precipitation-runoff relationships. I also thought it was a valuable addition to include a range-wide assessment to highlight not only the application of the approach but also to show how basin response to drought varies as a function of elevation.

> We agree that our results motivate improvements in ET parametrizations, especially since the observed shifts extend to the majority of water basins in the Sierra Nevada and occur regardless of predominant geology.

General comments: 1. The recent 2012-2015 drought period considered by the authors is inconsistent (P2 L33) with the official declaration period of drought on the website provided (P3 L25), which gives 2012-2016. The modelling exercises will need to be repeated to include water year 2016 if the authors would like to stick to the official declaration of drought. As they do note that the results are sensitive to the duration of drought episodes, it would be worth including (and comparing) the 2012-2015 (what could be called 'peak drought' conditions) and 2012-2016 (the 'official drought') periods.

> Changes: We included water year 2016 to all results in this manuscript. We did not add a specific discussion on this addition as this was considered a little out of scope for this study.

I think it would strengthen the paper to include some additional modeling studies of longer droughts, even if these are hypothetical. Many water agencies (and model experiments) use 'design' droughts based on a few known drought episodes lumped together or informed by past extended droughts. While I see equation 3 was applied by lumping results together on Page 6, I would like to see how a few iterations of a design drought of decadal (or longer) scale change (or do not change) the results.

> We agree that scenarios involving longer droughts can inform water management, and hope that our results can improve those studies. As to using scenarios in the current study, that would involve developing different data sets that would have less certainty in assessing model response. All results in this paper rely on measurements, including detecting shifts in precipitation vs. runoff, assessing the performance of the PRMS model during droughts and wet periods, and comparing estimated and modeled water-balance components. While paleoclimatic datasets suggest prolonged, multi-decadal droughts in California, it would thus

be challenging to generate the observational dataset we need to fully apply our methods. In addition, drought vs. non-drought conditions in California have a strong interannual character because of the quasiperiodicity of El Nino–Southern Oscillation [8], meaning that investigating these shorter time scales is functional to answering our research questions.

Changes: we added all points above to the manuscript **(see lines 24ff page 7)**.

2. PRMS is a fine modeling approach, but in regions with known groundwater/surface water interactions, such as the volcanic Almanor sub-basin, why was a groundwater model not also developed and coupled to PRMS (i.e., GSFLOW?). I understand this may not improve results, but without showing that it does not, it leaves the reader wondering if such an addition would improve results.

We agree that coupling a groundwater model to PRMS could shed further light on shifts in the basin water balance, and potentially precipitation-runoff relationships. While a coupled version of PRMS is available (GSFLOW), setting it up with the same level of spatial data and parameterization as it was done for PRMS would be a major new study, requiring an effort that goes well beyond the scope of the research presented in this manuscript. Lacking the data needed to describe groundwater flow in the basin with a high level of rigor, it is our assessment that the simplified simulation of groundwater processes in PRMS is appropriate to meet the aims of this study. In this simplified setting, PRMS and many other hydrologic models with similar process representations are currently being used by water and energy forecasters in California and elsewhere to predict water supply at various time scales. Highlighting that such hydrologic models are prone to performance drops during shifts in precipitation-runoff relationships thus addresses an important aspect of operational hydrology.

Changes: we added a brief discussion on this point in our paper **(see lines 13ff page 8)**.

3. P8 L5: The authors misconstrue the definition of 'warm snow drought', perhaps because the definition can be misleading if no thresholds in precipitation anomalies are specified. Though the 2012-2015 drought may have had greater precipitation than the 1977 drought, many of these years in both drought periods satisfied the dry snow drought conditions demonstrated in Hatchett and McEvoy (2018). Warm snow droughts should only be defined by years with near to above-normal precipitation and below average snowpack.

Changes: we removed any reference to snow droughts in the manuscript, since these processes were not the focus of the present study.

4. P8 L25: Snow/rain ratios should not be based upon a single point in time value of accumulated precipitation and snowpack at that time (April 1), as this neglects numerous factors that may be controlling the state of the snowpack on this date. For example, if a third of an above-average snowpack was lost during a warm, humid period in late March, the value on April 1 would not correctly represent the fact that otherwise a year had experienced above normal snow and perhaps above normal rain/snow ratios. Just a hypothetical example. I suggest calculating snow fraction over the course of the year and using the total precipitation estimated as snow divided by total precipitation (ending on Apr 1) as a more robust metric.

We considered to add the above metric and assessed that its computation would require a fairly large amount of assumptions since data reported in Figure 2 are monthly. Also, some of these precipitation data lack co-located air-temperature (and optionally relative-humidity) data to estimate phase partitioning between rain and snow.

Changes: we removed snow/rain ratios from Figure 2 in the manuscript. These only had an illustrative purpose and were not necessary to answer our research questions.

5. The procedure used to estimate ET is very interesting, but as this method has not yet been accepted by the scientific community (as noted by a submitted paper on P5 L4 and L6), I need to see comparisons of these results with some standard, easily implementable methods to calculate ET.

The procedure is a modification of the well-cited approach by [5], which has been used in multiple papers since then [6, 3, 1, 7]. The main change compared to [5] is the addition of precipitation as predictor, which both recognizes the responsiveness of ET to precipitation, independent of NDVI, and provides a lower error. Further details can be found in [12].

Changes: we added further details about the procedure used to estimate ET in the main text **(see lines 1ff page 6)**, and also point the reader to the recently published paper by [12] – see an author formatted, accepted version at `https://escholarship.org/uc/item/4kb947md`.

6. The concept of climate-driven ET elasticity is very cool. I would recommend a schematic figure be added to highlight this concept. The additional model runs using longer drought episodes (warm and dry versus cool and dry at decadal time scales) could play into making this figure more robust by helping constrain the temporal and climate condition sensitivity of the elasticity.

Changes: we added a schematic to the manuscript as suggested **(see lines 3ff page 15)**.

7. The map figures (Figure 1) are not up to publication standards. They need to be projected with latitude and longitude coordinates. The bins of precipitation are far too large and substantial precipitation variability is lost. The inset map needs to be projected and should show the entire west coast of North America (or at least the western United States). I suggest simply binning precipitation by 50 mm increments to better highlight topographic gradients. The geologic mapping appears to be hand drawn (and is of very poor quality) and needs to be markedly improved. The map should also show the locations of the stations used in the study since they are referenced in the main text.

Changes: we modified Figure 1 following these suggestions, with only two exceptions: (1) we kept a fairly small number of bins in the precipitation map because this allowed us to highlight the sharp transition between the wet, western side of the basin and the dry, eastern side; this is the main point we would like to convey there; (2) the geological map was removed and we pointed to publicly available reports where such maps are available [9].

Specific Comments: P3 L11: Please change all instances of Oroville Lake to the correct title, “Lake Oroville”

Changes: done.

P3 L26: When referring to climate, one must not neglect temperature if discussing precipitation. Please change to “dry, hot summers and wet, mild winters”.

Changes: done.

P3 L28: It would help to add a figure demonstrating the basin hypsometry to quantify 'most'. Does "most" mean 59% or 92%?

Changes: done **(see lines 16ff page 4)**.

P4 L4: Add 'anomalous' before 'low'

Changes: done.

P4 L13: Should be "Cascade Range".

Changes: done.

P4 L14: Suggest to add more geologic context, specifically on soils.

Changes: done **(see lines 1ff page 5)**.

P4 L22: Please provide the temporal resolutions used. Perhaps a table could be used?

Changes: We added a table in the Supporting Information detailing all datasets used, their temporal-spatial resolution, and their role in our study **(see Table S1)**.

P4 L22: What are the resolutions of the spatially distributed indices?

Changes: We added a table in the Supporting Information detailing all datasets used, their temporal-spatial resolution, and their role in our study **(see Table S1)**.

P4 L29: What types of precipitation gauges were used? Were gauges heated? Are there concerns for undercatch?

Changes: We added a paragraph in the manuscript to discuss all these points – thank you **(see lines 27ff page 5)**.

P5 L1: Which PRISM products were used, 4 km or 800 m? Monthly or daily?

Changes: We added a table in the Supporting Information detailing all datasets used, their temporal-spatial resolution, and their role in our study **(see Table S1)**. Same information was added in **Section 2.2** for completeness.

P5 L25: What alpha level was significance assessed at?

Changes: significance level $\alpha = 5\%$ **(see line 1 page 7)**.

P8 L1: I'm a bit confused here, is the paper referring to flow below Oroville, or total inflow to Oroville? If the former, these numbers need to be prefaced by discussion of water deliveries that may have been subject to changed allocations in response to the drought. Similarly, the storage value in the reservoir doesn't really add much, given that a reservoir's operational goals might be to completely drain the reservoir by the end of the water year. If

additional context is provided, then this number becomes meaningful. Last, please correct 'norm' to the proper spelling 'normal'.

Based on our understanding of the technical report by DWR [2], this is full-natural flow at Oroville, which is comparable to inflow to the reservoir and is independent from water-allocation decisions downstream of the dam. It is true that reservoir storage is highly seasonal, but it also responds to multiple objectives that may require reservoir level to be maintained to a certain high level (e.g., recreational reasons, hydropower production, multi-year carryover for water supply, etc).

Changes: no change needed, besides removing the word 'normal'.

P8 L3: For consistency with the discussion on the 1987-1992 drought on P8 L8, please include the temperature anomaly for the 2012-2016 (note I used the official period there, not 2012-2015). P8 L9: Section should be singular.

Changes: done.

P8 L9: The start of this paragraph is a bit strange, i.e., is there any reason to suspect that runoff seasonality should not be preserved? I think this sentence could be removed or replaced with some more insightful findings. Perhaps discuss any temporal shifts?

Changes: we revised the paragraph as suggested (**see lines 23ff page 10**).

P10 L11: Can you add some additional clarity about these winter precipitation events? I would expect these to be extreme precipitation rain-on-snow events to produce peak flows, but as it is written, any precip event could generate a peak flow event.

Changes: we revised the paragraph as suggested (**see lines 24ff page 11**).

P10 L30: I am balking at the use of 'observed' water balance, as that implies that the water balance has been completely observed, when in reality it is merely an estimate based upon models for precip (PRISM) and ET (NDVI-GAM approach). "Estimated" might be a better word but could confuse readers against "modeled". I leave this to the authors to ponder if a better descriptor could be used.

Changes: wherever possible, we used the word *estimated* to clarify that none of those water fluxes were directly measured, but are the results of statistical models. We also specified that in Section 2.3.3. and all results thereof, the word *observed* is used as opposed to PRMS-*modeled* fluxes (**see lines 24ff page 9**).

P11 L8: What is "soft data"?

Measuring sub-surface-storage decline is challenging, meaning one has to rely on indirect observations (e.g., magnitude of low flows or rate of seasonal flow from springs). Some of these indirect observations are discussed in [4]. This was our intended meaning of 'soft data'.

Changes: we revised the paragraph using some of the wording we used here to clarify this point (**see lines 17ff page 12**).

P12 L25: Suggest to add a citation for the second and third sources of uncertainty.

[1], which is cited at the very beginning of this paragraph, is the main citation for all the sources of uncertainty in this paragraph. No change needed.

P13 L6: I like the concept of tree mortality (despite the myriad complexities controlling tree mortality in addition to ET like pests, disease, etc), but I feel like this sentence detracts from the previous, powerful statement defining climate elasticity of ET. Ending the paragraph with this definition and instead another elucidating sentence about the value of this metric would be a strong way to close out this subsection.

Changes: we rephrased the paragraph as suggested.

Table 1: Please add the period of record means (or medians) for each variable. These statistics are not provided in Figure 2 as the caption implies.

Figure 2 reports annual average maximum and minimum temperature, and annual quartiles of cumulative precipitation and April-1 SWE. We aggregated these values to obtain statistics in Table 1.

Changes: caption in Table 1 was revised accordingly.

Reply to Reviewer #2

Overall, this is a very interesting and relevant piece of work. However, authors tend to generalize and leave the reader hanging in some cases which raises some questions. Please see below specific comments for your attention.

Title needs modification, it does not communicate the focus of the study, especially after reading the first sentence of the abstract.

> Our choice was to summarize the main take-away of this paper as title. Albeit infrequent, this choice is increasingly popular in hydrologic literature [13, 10, 11] and it does, in our opinion, communicate what is the main focus of the study, that is, how the buffered response of *ET* to precipitation variability impacts the water balance of Mediterranean mixed rain-snow catchments. At the same time, we acknowledge that some word choices in the current title may sound unfamiliar for the broad audience (e.g., precipitation-runoff relationships).
>
> Changes: we proposed a revised title.

Did authors focus on the effect of drought on shifts in precipitation-runoff relationships and the performance of the model? Authors need to clarify the catchment/s studies. Or they use one catchment with sub-basins, and if so, the results reported should be for the basins?? The abstract has to be summarized and comprehensive.

> We focused on three points: (1) we quantified shifts in precipitation-runoff relationships in four nested catchments of the Feather River; (2) we assessed performances of the PRMS model in these nested catchments during droughts and in particular during periods corresponding to shifts in the water balance; (3) by leveraging the fact that the performance of the model was sensitive to these shifts, we identified the water-balance term for which accuracy during droughts was statistically different from accuracy during wet periods (*ET*). We concluded that a different response time of *ET* to precipitation variability is the likely cause of these shifts. We called this climate elasticity of *ET*. This point is general and goes beyond the specific catchments we considered.
>
> Changes: we revised the Abstract to clarify these points.

Introduction

The motivation and novelty of the study is very week. The introduction lacks coherence. For example, the reader has to be able to identify:

1. overall effect of droughts on water balance in Med climate, with specific examples.

2. The approaches of examining precipitation-runoff relationships and how successful they have been in the same climate.

3. What has been done so far in relation to precipitation-runoff relationships or water balance studies within the same climate. This indicate the contribution of the work and its novelty.

> Changes: we completely revised the Introduction following the structure recommended by the Referee. We also rearranged the sub-sections in Discussion to better align with the

Introduction.

Research questions: Authors need to improve and rewrite all their research questions to be clear. It's very difficult for the reader to understand them. They raise a lot of questions: what is the water-balance predictive skill? Are the authors referring to the potential of the model to predict shifts during drought and non-drought periods???

Changes: we worked on improving clarity of our research questions.

What is the suitability of the generalized additive model in predicting ET for the catchment? Was it used before?

The procedure is a modification of the well-cited approach by [5], which has been used in multiple papers since then [6, 3, 1, 7]. The main change compared to [5] is the addition of precipitation as predictor, which both recognizes the responsiveness of ET to precipitation, independent of NDVI, and provides a lower error. Further details can be found in [12].

Changes: we added further details about the procedure used to estimate ET in the main text **(see lines 1ff page 6)**, and also point the reader to the recently published paper by [12] – see an author formatted, accepted version at `https://escholarship.org/uc/item/4kb947md`.

What were the characteristics of the Landsat-based annual-averaged NDVI? How was it derived? Is it a product or authors derived their own product through estimation? Any preprocessing of the image?

Landsat 5, 7 and 8 were used to map NDVI at 30-m resolution. Values were calculated from the Tier 1 surface-reflectance product downloaded from Google Earth Engine. NDVI values among different Landsat sensors were homogenized by cross-calibrating Landsat 7 (NDVI in 2012) and Landsat 8 (NDVI in 2013-2016) into Landsat 5. Annual Landsat NDVI maps were generated by averaging all pixels in a water year. Pixels with shadow, snow, or cloud were excluded from the calculation.

Changes: we thought this too much detail for the manuscript and point the reader to the recently published paper by [12] for firther details – see an author formatted, accepted version at `https://escholarship.org/uc/item/4kb947md`.

I suggest authors use their main objectives as subheadings for the reader to understand the findings and their implications.

The three subsections of the Discussion do not directly relate to the main objectives stated in the Introduction, but look at results from other perspectives to further clarify their implications.

Changes: the main results of the paper were clarified in the first paragraph of the discussion.

References

[1] R. C. Bales, M. L. Goulden, C. T. Hunsaker, M. H. Conklin, P. C. Hartsough, A. T. O'Geen, J. W. Hopmans, and M. Safeeq. Mechanisms controlling the impact of multi-year drought on mountain hydrology. *Scientific Reports*, 8:690, 2018.

[2] DWR. The 1976-1977 California drought - a review. Technical report, State of California, Department of Water Resources, 1978.

[3] A. W. Fellows and M. L. Goulden. Mapping and understanding dry season soil water drawdown by california montane vegetation. *Ecohydrology*, 10(1):e1772, 2017.

[4] G. J. Freeman. Climate Change and the Changing Water Balance for California's North Fork Feather River. In *Proceedings of the 79th Annual Western Snow Conference*, pages 71–82, 2011.

[5] M. L. Goulden, R. G. Anderson, R. C. Bales, A. E. Kelly, M. Meadows, and G. C. Winston. Evapotranspiration along an elevation gradient in California's Sierra Nevada. *Journal of Geophysical Research: Biogeosciences*, 117(G3), 2012.

[6] M. L. Goulden and R. C. Bales. Mountain runoff vulnerability to increased evapotranspiration with vegetation expansion. *Proceedings of the National Academy of Sciences*, 111(39):14071–14075, 2014.

[7] M. L. Goulden and R. C. Bales. California forest die-off linked to multi-year deep soil drying in 2012–2015 drought. *Nat. Geosci.*, 12:632–637, 2019.

[8] D. Griffin and K. J. Anchukaitis. How unusual is the 2012–2014 california drought? *Geophysical Research Letters*, 41(24):9017–9023, 2014.

[9] K. M. Koczot, A. E. Jeton, B. McGurk, and M. D. Dettinger. Precpitation-Runoff Processes in the Feather River Basin, Northeastern California, with Prospects for Streamflow Predictability, Water Years 1971-1997. Scientific Investigations Report 5202, USGS, 2004.

[10] T. H. Painter, M. G. Flanner, G. Kaser, B. Marzeion, R. A. VanCuren, and W. Abdalati. End of the little ice age in the alps forced by industrial black carbon. *Proceedings of the National Academy of Sciences*, 110(38):15216–15221, 2013.

[11] M. S. Raleigh and E. E. Small. Snowpack density modeling is the primary source of uncertainty when mapping basin-wide swe with lidar. *Geophysical Research Letters*, 44(8):3700–3709, 2017.

[12] J. W. Roche, Q. Ma, J. Rungee, and R. Bales. Evapotranspiration mapping for forest management in california's sierra nevada. *Front. For. Glob. Change*, 2020.

[13] N. Wever, C. Fierz, C. Mitterer, H. Hirashima, and M. Lehning. Solving Richards Equation for snow improves snowpack meltwater runoff estimations in detailed multi-layer snowpack model. *The Cryosphere*, 8(1):257–274, 2014.

~~Evapotranspiration feedbacks shift annual precipitation-runoff relationships~~ Climate elasticity of evapotranspiration during multi-year droughts ~~in a~~ shifts the water balance of Mediterranean ~~mixed~~ rain-snow ~~climate~~climates

Francesco Avanzi[1, 2], Joseph Rungee[3], Tessa Maurer[1], Roger Bales[3, 1], Qin Ma[3, 4], Steven Glaser[1], and Martha Conklin[3]

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, 94720, Berkeley, California, USA
[2]CIMA Research Foundation, via Armando Magliotto 2, 17100, Savona, Italy
[3]Sierra Nevada Research Institute, University of California, Merced, 95343, Merced, California, USA
[4]Department of Forestry, Mississipi State University, 39762, Mississippi State, Mississipi, USA

Correspondence: Francesco Avanzi (francesco.avanzi@cimafoundation.org)

Abstract.

Multi-year droughts in Mediterranean, rainfall-dominated climates lead to shifts in the water balance, the fundamental precipitation-allocation rule across runoff, evapotranspiration, and subsurface storage. Despite their fundamental implications, mechanisms causing these shifts as well as the extent to which they impact mixed rain-snow regimes remain largely unknown and are not well represented in hydrologic models. Focusing on the headwaters of the California's Feather River, we ~~investigated how multi-year droughts affect the water balance of Mediterranean~~ found that in these mixed rain-snow ~~catchments. Droughts in these catchments saw~~ Mediterranean catchments a lower fraction of precipitation was allocated to runoff during multi-year droughts compared to non-drought years. ~~This shift in precipitation-runoff relationship was larger in a~~ Comparing surface-runoff-dominated ~~than in a~~ and subsurface-flow-dominated ~~catchment —~~ sub-basins, we observed different responses to drought: 39% ~~and~~ lower runoff during drought in the surface-runoff-dominated versus 18% less runoff ~~, respectively, for a representative precipitation amount. The performance~~ in the subsurface-flow-dominated sub-basin. The predictive skill of the PRMS hydrologic model in these catchments decreased during droughts, ~~particularly those causing larger shifts in the annual precipitation-runoff relationship. Evapotranspiration~~ with evapotranspiration (ET) ~~was~~ being the only water-balance component ~~for which predictive accuracy~~ besides runoff for which this drop in performance during drought vs. non-drought years was ~~consistently different. Besides a systematic bias during all years~~statistically significant. In particular, the model ~~tended to relatively overestimate drought~~ underestimated the buffered response of ET ~~and to underestimate non-drought~~ to interannual climate variability, or climate elasticity of ET. ~~Modeling errors for~~ Differences between simulated and data-driven estimates of ET ~~during droughts were somewhat correlated with maximum and minimum annual temperature as well as~~ were well correlated with accompanying data-driven estimates of changes in sub-surface storage (ΔS, r = ~~-0.45, -0.57, and 0.23, respectively). These correlations point to the interannual response of ET to climate , or climate~~ 0.78). This correlation points to shifts in precipitation-runoff relationship being evidence of a hysteretic response of the water budget to climate elastic-

ity of ET ~~, as the likely driver of the observed shifts in precipitation-runoff relationship during droughts in Mediterranean mixed rain-snow regions; underestimation of this response caused increased modeling inaccuracy during droughts. Improved predictions of interannual variability of~~ during and after multi-year droughts. This hysteresis is caused by carryover storage offsetting precipitation deficits during the initial drought period, followed by vegetation mortality when storage is depleted and subsequent post-drought vegetation expansion. Our results point to a general improvement in hydrologic predictions across drought and recovery cycles by including the climate elasticity of ET~~are necessary to support water-supply management in a warming climate and could be achieved~~, and better accounting for actual subsurface water storage in not only soil, but deeper regolith that also stores root-accessible water. This can be done by explicitly parametrizing ~~feedback mechanisms across~~ carryover storage and feedback mechanisms capturing vegetation response to atmospheric demand for moisture~~, ET, and multi-year carryover of subsurface storage~~.

Copyright statement. TEXT

1 Introduction

~~Droughts have a profound impact on ecosystems and societies (?), especially because they~~

Regions with a Mediterranean climate receive the bulk of precipitation during winter, while summers are dry (**?**). This seasonal imbalance in precipitation distribution, coupled with asynchronicity between precipitation input and potential-evapotranspiration demand (**???**), complicates understanding and management of multi-year droughts and their impact on water supply (**?**). Water supply is the output of a water balance, that is, $Q = P - ET - \Delta S$, where Q is runoff, P is precipitation, ET is evapotranspiration, and ΔS is the change in sub-surface storage. Major droughts reduce P, which directly reduces Q, but quantifying this non-linear impact is complicated by the additional, and often-overlooked, effects of drought on ET and ΔS. Further understanding the impact of droughts on the water balance of Mediterranean climates is relevant to ecosystem services and water security (**?**), especially because droughts can be more persistent than other water risks (**??**) ~~. Some examples include the Dust Bowl drought in the 1930s, which triggered long-lasting changes in the economy and population distribution of the United States (?), the Millennium Drought in south-eastern Australia (1997-2009, see ??), the persistently dry conditions across the U. S. southwest in the early 2000s (?), the large-scale European drought in the 1920s (?), and the 2016 El-Niño drought in South Africa (?). Since aridity will increase in frequency and extent under a warming climate (??), understanding the impact of droughts on the hydrologic budget is relevant to water supply, ecosystem services, and water security (?).~~ and may increase in frequency and geographic extent under a warming climate (**??**).

~~Precipitation deficit has frequently been assumed as the primary proxy to explain changes in water supply during droughts (?). While low-precipitation periods generally lead to a decrease in runoff, other factors can exacerbate or alleviate this response. Some of these factors include concurrent air temperature (??),~~ Five possible mechanisms may alter ET and ΔS during and after drought and thus intensify or alleviate the impact of changing P on Q in Mediterranean climates (**?**). These mechanisms are seldom measured and remain incompletely understood, and are thus not well represented in hydrologic modeling (**?**) . First, ET may not increase or reduce in proportion to P, thus shifting the fraction of local P partitioned to Q during dry periods. Direct measurements of evapotranspiration in non-drought years show a relatively muted response of ET to P (**?**). Second is priority allocation of P to ET, facilitated in some Mediterranean climates by slow-draining soils and thus ample water for dry-season ET (**??**). For example, the importance of water stored in deeply weathered granitic rock in the Sierra Nevada to the productivity and survival of forest trees during the summer dry season is well established (**??**). A set of coastal California catchments with mixed-coniferous-broadleaf evergreen forests or deciduous oak savanna, and with low water storage, showed limited interannual variability in ET, despite variable interannual precipitation (**?**). This is in contrast to similar precipitation-limited sites with deeper regolith storage in the Sierra Nevada that showed significant response to interannual precipitation amounts (**?**). Third, pre-drought ~~aridity (**?**), tree mortality and evapotranspiration (**?**), sub-surface storage (**???**),~~ aridity (**?**), or the amount of storage and carry-over from previous years (**???**), can affect the amount of stored water available for dry-season evapotranspiration, and ~~geology (e.g., granitic vs. volcanic bedrock, see **???**). Current drought indices do consider a broad spectrum of climatic variables (**??**), but catchment properties can still challenge drought-impact assessments (**?**). In the African Sahel, for example, a multi-decadal drought has led to a poorly understood increase in runoff (*Sahelian paradox*, see **?**).~~ thus the relation between P and Q. The depletion of this multi-year storage was the major driver for the moisture stress that led to widespread tree mortality in the Sierra Nevada in 2015, during the multi-year drought (**?**). Fourth, changes in either meteorology and evaporative demand (**??**) or vegetation structure and transpiration can alter ET and either increase or decrease Q. Evaporative demand in the Sierra Nevada, indicated by moisture stress, during the 2012-16 drought was higher than in the 1987-92 drought, owing to the 1-2°C warmer temperature (**?**). Variations in vegetation structure associated with drought-induced mortality and wildfire changed transpiration in the southern Sierra during the 2012-16 drought, and thus the $P-Q$ relation (**?**). Fifth, the spatial heterogeneity and the covariance between P, ET, and ΔS shift the location and relative importance of source regions for Q across a catchment (**?**). That is, interannual differences in spatial patterns of P, ET, and available subsurface water storage elicit a non-linear response in Q. For example, higher elevations of the Sierra Nevada have $P > ET$ even in dry years, whereas lower elevations can switch from having $P > ET$ in wet years to $P < ET$ in dry years. These differences can be driven by geology (see e.g. **???**) as well as climate and the interaction of the two over long time periods (**?**).

~~An inconsistent response of runoff to droughts may be evidence that these periods lead to changes in catchment functioning, similarly to other catchment-climate coevolution processes (**??**). Such changes~~ Changes in runoff response during droughts have been observed in Australia (**?**), California (**?**), and China (**?**) and have ~~usually~~ been quantified as statistical shifts in the precipitation-runoff relationship, i.e., an empirical regression between annual precipitation and annual runoff (**???**). ~~While these lumped interpretations~~ These lumped approaches allow one to predict the occurrence of shifts based on catchment and drought characteristics (**?**), ~~the internal catchment mechanism behind themhas not yet been fully clarified. Runoff is ultimately the output of a water balance, that is, $Q = P - ET - \Delta S$, where Q is runoff, P is precipitation, ET is evapotranspiration, and ΔS is the change in sub-surface storage. In a water balance, shifts thus correspond to a different allocation of P~~ but they do not shed light on how the five internal catchment mechanisms outlined above interact to cause them. Since performance of hydrologic models appear to degrade when these shifts occur (**??**), this knowledge gap limits improvements in hydrologic-model predictive accuracy and challenges the identification of drought-management solutions (**?**). Moreover, shifts in the precipitation-runoff relationship have been quantitatively characterized only in snow-free regions (**??**): given the mitigating and delaying effect of snowmelt recharge on soil-water drawdown (**?**), it is currently unclear whether shifts in the water balance could occur even in Mediterranean catchments that are seasonally covered by snow (**?**). Unraveling interactions across ~~ET, ΔS, and Q between drought and non-drought periods. Unraveling the interplay across~~ water-balance components ~~is a key to clarify~~ (P, ET, ΔS, and Q) is key to clarifying the mechanisms behind shifts in the precipitation-runoff relationship~~during droughts (**?**).~~ , reach a better understanding of the water balance during droughts, and ultimately inform better water-management decisions.

~~A water-balance perspective of droughts is essential in a Mediterranean climate, where precipitation is concentrated in winter and summers are dry (**?**). In these regions, water stored in the form of snow or in the regolith can support evapotranspiration during multi-year droughtsand offset precipitation deficit at the expenses of runoff (**????**), a mechanism that is further exacerbated by increasing temperatures and thus increased aridity (**?**).~~ California, a mixed rain-snow region with a ~~markedly~~ Mediterranean climate, has seen four officially designated droughts since the 1970s ~~(water years 1976-77, 1987-92, 2007-09, and 2012-15, see **?**)~~ (water years 1976-77, 1987-92, 2007-09, and 2012-16, see **?**). Because most ~~of~~ precipitation in the state falls in the north and during winter, water supply is mostly generated in ~~mixed rain-snow,~~ geologically and topographically complex headwaters, while water is mainly consumed in lowland regions ~~.~~ further south. Rising temperatures are threatening this equilibrium (**??**), ~~but the impact of droughts on Mediterranean, mixed rain-snow catchments has rarely been studied from a water-balance perspective, meaning both hydrologic-model predictive accuracy and drought-management solutions are still uncertain (**?**).~~

~~The~~ meaning the four Californian droughts between the 1970s and 2010s offer ~~an~~ a decision-relevant opportunity to clarify the mechanisms behind shifts in precipitation-runoff relationships in a Mediterranean, mixed-rain snow climate ~~,~~ as well as the adequacy of hydrologic models to simulate them. To achieve this goal, we used detailed water-balance indices and hydrologic modeling ~~(PRMS, see **??**) to~~ (PRMS, see **???**) to address three research questions~~: First, what shifts~~ . First, do droughts cause shifts in the precipitation-runoff relationship of mixed rain-snow catchments in a Mediterranean climate~~? Second, do~~ , similar to shifts in rainfall-dominated regions? If so, how is their occurrence influenced by predominant geology? Second, does the occurrence of these shifts affect ~~water-balance predictive skill in basins with different predominant geology~~the predictive accuracy of hydrologic models? Third, what ~~is the catchment mechanism~~ catchment mechanisms are causing shifts during ~~droughts as opposed to wet periods~~ drought versus wet periods in Mediterranean regions?

2 Methods

We focused on the Feather River upstream of ~~Oroville Lake~~ Lake Oroville in the Sierra Nevada of California (∼9300 km^2, see Figure 1) and on three of its sub-basins with contrasting geology (see Section 2.1 for details). Water from the Feather is both exploited locally for hydropower production by Pacific Gas & Electric (PG&E, see **?**) and impounded by Oroville Dam to support water supply across the state through the State Water Project (**?**).

Our research followed three main steps (Sections 2.3.1 to 2.3.3). First, we quantified shifts during droughts in the observed precipitation-runoff relationship of the three (sub-)basins with serially complete full-natural-flow data (see Section 2.2 for details about full-natural-flow data). Second, we assessed the performance of the PRMS hydrologic model in predicting full-natural flow in all (sub-)basins~~and~~ , in particular during droughts, in order to gain insight into the potential impact of these periods on predictive accuracy. Third, we identified the driver of PRMS predictive accuracy during droughts and its potential relationship with shifts in the precipitation-runoff relationship by comparing ~~observed and simulated~~ independently estimated and PRMS-simulated basin-wide mass-balance ~~indexes~~ indices ($P, ET, \Delta S, Q$) in the (sub-)basins with serially complete full-natural-flow data.

We focused on water years 1970 to ~~2015 due to both data availability and data quality~~2016 in order to cover the complete timespan of the most recent four multi-year droughts in California. The water year is defined as October 1st to September 30th and it is indicated with the calendar year in which it ends (e.g., water year 2015 went from October 1st, 2014 to September 30, 2015). We defined drought water years following the official declarations of the State of California: 1976-77, 1987-92, 2007-09, and ~~2012-15~~ 2012-16 (see https://water.ca.gov/Water-Basics/Drought, visited July 19, 2019).

2.1 Study area

The climate of the Feather is Mediterranean, with dry, hot summers and wet, mild winters. Elevation ranges from ∼250 m above sea level (ASL) at Oroville Dam (the outlet of the basin) to ∼2900 m ASL at Mt. Lassen, but ~~most~~ approximately 90% of the catchment lies below 2000 m ASL ~~(**?**). Mixed~~ (see Figure 1(c)). Therefore, mixed rain-snow and rain-on-snow events are frequent across the basin ~~(**?**)~~(**??**). The water balance of the Feather has experienced recent warming-related changes, including declining runoff and peak snow accumulation (**??**), forest growth (**?**), and a rise in the rain-snow transition ~~line~~ elevation (**?**).

The Feather is the most northern of the thirteen basins draining from the Sierra Nevada into the Sacramento-San Joaquin valley (see **?**). Contrary to most of these catchments (**??**), some headwaters of the Feather lie in the eastern, rain-shadowed side of the Sierra divide (Figure 1~~)~~(b) and (d)): mean annual precipitation thus ranges from ∼2800 mm in the western side of the basin to less than 800 mm in the eastern side. Because anomalously low precipitation has been suggested as a key predictor of shifts during droughts (**?**), this basin is an ideal case study to answer our questions.

Our study considered two spatial scales (Figure 1): the Feather at Oroville Dam and three of its headwater sub-basins: Almanor (∼1100 km^2, 1400-2900 m ASL, rain-shadowed), East Branch (∼2600 km^2, 725-2550 m ASL, rain-shadowed), and Middle Fork (∼2700 km^2, 480-2660 m ASL, partially rain-shadowed). Hydrologic studies on the Feather River at Oroville are abundant (see for example **????**, and references therein), whereas headwater sub-basins have rarely been studied as stand-alone catchments ~~(see examples in **???**)~~(see examples in **????**).

The Almanor sub-basin lies at the intersection between the granitic Sierra Nevada and the volcanic Cascade Range and is thus dominated by a porous, volcanic bedrock~~(see Figure 1).~~ , with soil texture being predominantly composed by silt (see a geologic and soil map in **?**, , page 12) This sub-basin is ~~mostly~~ largely fed by subsurface flow (**?**) and has exhibited a 30% decline in water-year inflow to ~~Almanor Lake~~ Lake Almanor (located at the outlet of this sub-basin) since the 1960s. This drop is attributed to missed groundwater-recharge opportunities due to decreasing snow accumulation (**?**). The geology of the East Branch and the Middle Fork includes impervious granitic outcrops similar to the rest of the Sierra Nevada and an alternation between sand and silt. Water supply in these two sub-basins is dominated by surface runoff (?), but the East Branch is significantly drier than the Middle Fork because it is fully rain shadowed.

2.2 Data

Data include daily full-natural flow at the outlet of the four (sub-)basins under study; in-situ precipitation, air temperature, and snow water equivalent (SWE) at various temporal resolutions; and estimated annual spatially distributed water-balance indices of precipitation, evapotranspiration, and variation in sub-surface storage. Table S1 in the Supporting Information further describes all data used, their spatial and temporal resolution, whether they were measured/estimated in situ or remotely, and their use in the paper.

Full-natural (~~or~~ unimpaired) flow is a mass-balance reconstruction of water supply as if no dam or other man-made infrastructure affected it (?). For the Almanor and East-Branch sub-basins, these data were provided by Pacific Gas & Electric (PG&E) at a daily resolution for water years 1970 through 2017. For the Middle Fork sub-basin and for the Feather River at Oroville, data were obtained from the California Department of Water Resources (DWR) at daily ~~and monthly resolution , respectively~~ resolution (water years 1987 to 2018 and 1985 to 2018, respectively).

In-situ daily precipitation from ten stations and daily maximum and minimum air temperature from three stations across the Feather river basin were obtained from PG&E, which routinely uses them as input for the PRMS hydrologic-forecasting model ~~(see details in ?)~~(see location of these stations in Figure 1 and ?, for further details). These data were quality checked and serially ~~gap-filled~~ gap filled by the company (water years 1970 to 2017). Additional data of monthly in-situ precipitation and manual snow water equivalent were downloaded from the California Data Exchange Center (https://cdec.water.ca.gov/, visited July 19, 2019) to study drought characteristics across the Feather (see ~~Table S1-S2 in the Supporting Information).~~ a complete list and their location in Table S2-S3 and in Figures S1 and S2, respectively).

Precipitation gauges used in the Feather river basin are managed by various operators, with the California Department of Water Resources collecting and archiving the data (Table S2). The design of these sensors resembles the one in use by the SNOTEL network throughout the western US (https://www.wcc.nrcs.usda.gov/about/mon_automate.html, visited February 1, 2020). Most gauges are unheated and some are manual; most are also located in small clearings where wind speed is low, which suggests that wind-driven undercatch is locally low, especially below the seasonal rain-snow line (?). Because precipitation gauges considered in this study are predominantly located in forested valleys below 2000 m ASL (see Figures 1 and S1) and many of them were not colocated with wind-speed measurements, we did not correct for undercatch (?). Undercatch and plugging increase in snow-dominated areas (??).

Spatially distributed, estimated annual precipitation (P) was obtained by accumulating daily 800-m maps from the Parameter-elevation Relationships on Independent Slopes Model (PRISM, see ?). Spatially distributed annual ET was estimated ~~using a generalized additive model (GAM) between single-term power-function transformations of~~ by extending the Landsat calibration approach by ?? to include the average of the current and previous year's precipitation as predictor in addition to Landsat-based annual-averaged NDVI (Normalized Difference Vegetation Index, 30 m)~~and the average of the current and previous year's~~

~~precipitation (**?**). The~~ . Both predictors were used as single-term ~~power transformations~~ power-function transformations that were developed by individually regressing the NDVI and PRISM data with 13 flux-tower measurements of evapotranspiration ~~(**?**). Variation in~~ in California (see a list and a map in **?**). Performance of this approach was evaluated by removing an individual water year for model building and then evaluating on the water year removed. Results showed an improved fit to calibration data for wet sites compared to only using NDVI. Developing this ET product was the scope of earlier research and we refer to **?** for further details. Estimated variations in annual basin-wide subsurface storage was estimated as the residuals of $P-ET-Q$, where P is basin-wide mean-annual PRISM-based precipitation, ET is basin-wide mean-annual ~~GAM-estimated~~ estimated evapotranspiration, and Q is annual full-natural flow. Landsat-based data were available for water years 1985-2016; PRISM maps were processed for the same period. PRISM data have a pixel size of 800 m, which we downscaled to 30 m using a nearest-neighbor approach to match that of Landsat.

2.3 Analyses

2.3.1 Shift in precipitation - runoff relationship

We detected shifts in the precipitation-runoff relationship by fitting a multivariate regression across annual cumulative full-natural flow (target variable), basin-wide annual precipitation, and a categorical variable denoting drought and non-drought years (**??**):

$$Q_{\mathrm{BC}} = b_0 + b_1 I + b_2 P + \epsilon, \tag{1}$$

where I is a categorical drought variable (1 for drought years and 0 for non-drought years), b_0, b_1, b_2 are regression coefficients, ϵ is noise, and Q_{BC} is annual full-natural flow transformed according to a Box-Cox transformation following the arguments in **???**:

$$Q_{\mathrm{BC}} = \frac{Q^{\lambda}-1}{\lambda}. \tag{2}$$

While λ should in principle be estimated from data to ensure linearity and heteroscedasticity (**?**), we assumed $\lambda = 0.25$ as suggested by **?** and references therein.

If different from zero, parameter b_1 indicated a shift of the precipitation-runoff relationship during droughts. This parameter is usually negative, as shifts during droughts tend to decrease runoff compared to precipitation deficit alone (**??**). We assessed the statistical significance of coefficient b_1 and concluded that the shift during droughts was statistically significant if the sign of the confidence bounds agreed ~~(**?**)~~(significance level $\alpha = 5\%$, see **?**). We performed this analysis for the Feather River at Oroville (~~1985-2015~~1985-2016) and the Almanor and East-Branch sub-basins (~~1970-2015~~1970-2016), for which we had serially complete time-series of annual full-natural flow.

Rather than directly using PRISM maps to estimate basin-wide precipitation, we tilted their monthly mean surfaces using precipitation data at the ten in-situ stations available to this study (see again Section 2.2). This operational procedure (called

DRAPER) is routinely used by PG&E forecasters on the river to force PRMS and is believed to provide more representative precipitation patterns for this basin than simply using PRISM surfaces (see **??**, for details about the DRAPER algorithm).

We estimated the relative magnitude of the shift in precipitation vs. runoff (M_Q) for each (sub-)basin with serially complete time-series of annual full-natural flow by using the approach suggested by **?**:

$$M_Q = \frac{Q_{\mathrm{dry},\mathrm{P_I}} - Q_{\mathrm{dry},\mathrm{P}}}{Q_{\mathrm{dry},\mathrm{P}}}, \tag{3}$$

where $Q_{\mathrm{dry},\mathrm{P_I}}$ is the (predicted annual) full-natural flow for a representative (annual) precipitation during dry periods according to the shifted precipitation-runoff relationship (Equation 1, $I = 1$), while $Q_{\mathrm{dry},\mathrm{P}}$ is the full-natural flow for the same precipitation according to the non-shifted relationship (Equation 1, $I = 0$). We assumed as representative annual precipitation the mean between average and minimum annual precipitation across the entire period of record (see more details, including a schematic, in **?**). Here again, we used DRAPER to estimate this representative precipitation, while full-natural flow in Equation 3 was not transformed.

? originally proposed M_Q to quantify the impact of the decade-long Millennium drought in south-eastern Australia ($\sim$1997-2009). The four Californian droughts under study were significantly shorter, so we applied Equation 3 by aggregating all drought years in one sample. We also quantified shifts for single droughts (m_Q) by assuming $Q_{\mathrm{dry},\mathrm{P_I}}$ to be the observed, average annual full-natural flow across each drought, and $Q_{\mathrm{dry},\mathrm{P}}$ to be the expected annual full-natural flow according to the non-shifted precipitation-runoff relationship (Equation 1, $I = 0$) and a reference annual precipitation equal to the average across each drought.

This shift-detection approach requires time series of annual precipitation and runoff. These time series are comparatively short in California, where for example water-supply forecasts started in the 1930s (**?**). Paleoclimatic datasets suggest prolonged, multi-decadal droughts in this region (megadroughts, see **?**), but investigating such scenarios would involve developing data sets that would have less certainty in assessing model response than the ground-based measurements used in this study. In addition, drought vs. non-drought conditions in California have a strong interannual character because of the quasiperiodicity of El Niño–Southern Oscillation (**?**), meaning that the time scale considered in this study is appropriate to answer our research questions.

2.3.2 PRMS performance during droughts: ~~flow~~streamflow

PRMS is a ~~semi-distributed~~ hydrologic model that solves mass and energy conservation across a given basin by discretizing it into Hydrologic Response Units (HRUs), regions of the basin that are assumed homogeneous (**?**). The model requires, as a minimum, inputs of daily precipitation and maximum-minimum temperature at one location, from which these data can be distributed to the grid of HRU centroids (**?**). In the Feather River PRMS model, ~~air temperature~~ air-temperature data from three stations are distributed using monthly lapse rates. Precipitation is distributed using the DRAPER method as outlined in Section 2.3.1 (**??**). Processes simulated include precipitation-phase partitioning, precipitation interception and storage by canopy, evapotranspiration, radiation distribution, snow accumulation and melt, infiltration and surface runoff, interflow, groundwater storage, and baseflow.

PRMS was calibrated and evaluated over the Feather River in the early 2000s by mostly focusing on full-natural-flow data between 1971 and 1997 (see **?**, for more details, including specific modules used by the model). While PRMS has been updated since then (the current version is 5 – June 2019), the model is currently set up for this river in version 2. The main differences between more recent versions and version 2 are the sub-surface components: version 2 separates sub-surface storage into superficial soil (including the recharge zone), a deeper sub-surface reservoir, and groundwater (**?**); more recent versions of PRMS consider instead a process-based separation into capillary, preferential, and gravity storage in addition to groundwater (**?**). ~~For this study, the representation of sub-surface~~ Also, PRMS has been recently coupled with a full groundwater model (GSFLOW, see **?**). Lacking the data needed to describe groundwater flow in the basin with a higher level of rigor, we assessed that the simplified simulation of groundwater processes in PRMS ~~version~~ 2 ~~was assumed to be sufficiently~~ is appropriate to meet the aims of this study, especially since it is representative of many conceptual models~~: for~~. For example, this version was implemented in inter-comparison tools like the Framework for Understanding Modeling Errors (FUSE, see **?**).

PRMS performance for full-natural flow was quantified using three different metrics: water-year Nash-Sutcliffe Efficiency (NSE), annual relative bias (relative to observations), and observed vs. simulated climate elasticity of streamflow. Because full-natural flow is prone to large errors, we smoothed the data and simulations using a five-day window before computing performance metrics.

NSE benchmarks the squared errors of simulations of a target variable (in our case, daily full-natural flow for each water year) against those obtained by using a long-term mean (**?**). The choice of this “long-term mean” can yield very sensitive results (**?**). In the Feather River basin, full-natural flow shows a large inter- and intra-annual variability (see some examples in **?**), implying that a mean across all water years would be a particularly poor benchmark (resulting in overoptimistic NSE values). On the other hand, a water-year mean would be an excellent predictor during dry years and a very poor predictor during wet years. In order to limit these spurious results, we benchmarked PRMS using a mean across all years from the same type according to the classification used by PG&E (henceforth, NSE$_{\mathrm{yrt}}$; see **?**, for some context on year types in California). We also computed the Nash-Sutcliffe Efficiency using log-transformed values (LogNSE$_{\mathrm{yrt}}$) as this metric is more sensitive to low flow (**?**).

Climate elasticity measures the sensitivity of annual streamflow (in this case, full-natural flow) to changes in a relevant climate variable, usually precipitation and potential evapotranspiration (**??**). Compared to NSE and bias, contrasting observed and simulated elasticity quantifies the performance of a hydrologic model in simulating inter-annual variability in streamflow and its relation to external forcings. Here, we considered absolute elasticity (nondimensional): $e_{Q/P}$ and $e_{Q/PET}$ are absolute elasticity to precipitation and potential evapotranspiration, respectively. Elasticity for both simulated and observed full-natural flow was computed in a bivariate linear framework using the approach proposed by **?**. As an independent benchmark, we also computed theoretical elasticity based on the Turc-Mezentsez formula (see again **?**). Similarly to shifts in precipitation-runoff relationships (see Section 2.3.1), we computed elasticity for the (sub-)basins with serially complete ~~time-series~~ time series of annual full-natural flow, that is, the Feather River at Oroville (~~1985-2015~~1985-2016) and the Almanor and East-Branch sub-basins (~~1970-2015~~1970-2016). We again used DRAPER to estimate basin-wide precipitation; potential evapotranspiration was estimated using the Jensen-Haise approach in PRMS (**?**).

2.3.3 PRMS performance during droughts: water balance

We quantified the performance of PRMS for the four components of the annual water balance by adapting Equation 1 to fit a regression between observed and simulated water-balance components during drought and non-drought years (period ~~1985-2015~~1985-2016, see Section 2.2 for data availability of Landsat data products):

$$P_{\text{obs}} = b_{0,P} + b_{1,P} I + b_{2,P} P_{\text{sim}} + \epsilon_P \tag{4a}$$

$$ET_{\text{obs}} = b_{0,ET} + b_{1,ET} I + b_{2,ET} ET_{\text{sim}} + \epsilon_{ET} \tag{4b}$$

$$\Delta S_{\text{obs}} = b_{0,\Delta S} + b_{1,\Delta S} I + b_{2,\Delta S} \Delta S_{\text{sim}} + \epsilon_{\Delta S} \tag{4c}$$

$$Q_{\text{obs}} = b_{0,Q} + b_{1,Q} I + b_{2,Q} Q_{\text{sim}} + \epsilon_Q, \tag{4d}$$

where, for example, P_{obs} and P_{sim} are observed and simulated basin-wide annual precipitation, while $b_{0,P}$, $b_{1,P}$, and $b_{2,P}$ are regression coefficients, respectively. If $b_{1,P}$ was statistically different from zero, then PRMS performance for precipitation during droughts was statistically different from that during non-drought years $(\alpha = 5\%)$. Similar definitions apply to the other water-balance components in Equation 4. This analysis was carried out for the (sub-)basins with serially complete time-series of annual full-natural flow (see Section 2.3.1).

Observed precipitation was assumed equal to the original PRISM maps in absence of other independent, distributed sources. Observed ET was ~~derived from the GAM-estimated maps~~ estimated through the product introduced in Section 2.2, while observed ~~ΔS~~ ΔS was estimated closing the water balance with observed full-natural flow (see again Section 2.2). Note that the word *observed* is used here in a statistical framework and in contrast to PRMS simulations to highlight that PRISM, the ET product, and full-natural flow are derived from measurements; however, we stress that none of them were measured directly, but are estimates of statistical models.

In addition to the standard four balance terms in Equation 4, PRMS includes a groundwater sink. This term is used to account for (often unknown) water losses in the sub-surface portion of the water balance. Because this sink together with ET represents the only internal water loss in the model, it was summed to ET_{sim} to fit a regression with observed evapotranpiration (note that this groundwater sink was set to zero in the original calibration of the Almanor sub-basin).

3 Results

3.1 Drought hydro-climatic summary

Average minimum daily air temperature at the three index stations of Canyon Dam (1390 m ASL), Quincy (1066 m ASL), and Bucks Creek Powerhouse (536 m ASL) showed a statistically significant increasing trend between water years 1970 and ~~2015~~

2016 (Kendall τ = ~~0.5~~0.54, p-value $<$ 0.01, α = 0.05, Sen's slope = ~~0.0448 ± 0.0134~~0.0477°C yr^{-1}, that is, $\sim$~~2~~2.1°C in 45 years, Figure 2). Neither average maximum daily air temperature nor annual precipitation presented a statistically significant trend (p-value = ~~0.57 and 0.99~~0.67 and 0.82, respectively – statistics for precipitation refer to median values across all stations, see Figure 2). April-1st SWE ~~also~~ had no significant trend for α = 0.05, but its p-value was significantly smaller (~~0.09~~0.05, statistics again refer to median values across all stations, see Figure 2). ~~The ratio between median April 1st SWE and median annual precipitation showed a statistically significant, yet slight, shift from snow to rain (Kendall τ = -0.22, p-value = 0.0279, α = 0.05, Sen's slope = -0.0055 yr^{-1}).~~

The four droughts under study had very different hydro-climatic characteristics (Figure 2 and Table 1). The 1976-77 drought was both the coolest (in terms of minimum temperature) and driest in our record ($\sim$~~56~~51% of average annual precipitation compared to ~~2012-15~~2012-16), and as a result 1976 and 1977 were the fourth and first driest water years in the State's historical record at that time (**?**). ~~Gauged flow from the Feather~~ Full-natural flow at Oroville was 43% and 24% of (contemporary) average for 1976 and 1977, respectively, the latter being a new record. ~~Storage of Oroville Lake on October 1, 1977 was $\sim$37% of the norm (?).~~

At the other end of the spectrum, the ~~2012-15~~ 2012-16 drought was the warmest ~~and (together with the short 2007-09 drought)~~ (in terms of minimum temperature) and the wettest during our study period (together with the 2007-09 drought, Figure 2 and Table 1). As a result, average April-1st SWE ~~/P~~ significantly declined during this drought compared to the ~~other three (0.26 as opposed to $\sim$ 0.50). A condition of comparatively high precipitation but lacking snow storage has been recently defined a warm snow drought(?)~~similarly wet 2007-09 drought. In between, the 1987-92 drought was the longest one (6 years), with five years classified as critically dry and one (1989) as dry (**?**). Average minimum temperature during ~~this drought~~ the 1987-92, 2007-09, and 2012-16 was $\sim$0.89 °C, $\sim$1.36 °C, and $\sim$1.98 °C higher than the 1976-77 drought, respectively.

~~While all droughts decreased annual water supply (see Sections 3.2), runoff seasonality was generally preserved between drought and non-drought years (Figure 3)~~Runoff timing from all (sub-)basins under study was highly seasonal, with peak flow occurring during winter and spring due to precipitation and snowmelt, and low flow occurring during the dry summer-fall season ~~. While the volcanic, subsurface-flow-fed Almanor sub-basin and the~~ (Figure 3). The surface-runoff-dominated East-Branch sub-basin ~~displayed comparable peaks in full-natural flow during winter, the latter~~ had a significantly lower flow during summer than the ~~former~~volcanic, subsurface-flow-fed Almanor sub-basin. This higher summer flow in the volcanic Almanor sub-basin compared to the granitic East Branch was consistent between drought and non-drought years.

3.2 Shift in precipitation vs. runoff

The four droughts under study caused a shift in the precipitation-runoff relationship for both of the two headwater sub-basins with complete annual data (Almanor and the East Branch) and the Feather River at Oroville (Figure 4). This shift means that the decrease in runoff observed during droughts in these (sub-)basins was larger than what could be explained by precipitation deficit alone (see Section 2.3.1 for details about the definition of shift). The 95% confidence bounds for b_1 were ~~-1.51 and -0.34; -2.29 and -1.00; and -1.45 and -0.29~~-1.49 and -0.38; -2.25 and -1.04; and -1.43 and -0.34, respectively (see Equation 1 for a definition of b_1), implying that this shift was statistically significant for all (sub-)basins. The magnitude of the shift (calculated using Equation 3) was -18%, ~~-39~~-38%, and -18% compared to precipitation allocation to runoff during non-drought years, respectively. Runoff from the volcanic Almanor sub-basin was thus more resilient to shifts during droughts than that from the East Branch, even if shifts were significant in all the (sub-)basins investigated.

The magnitude of these shifts varied from drought to drought and across (sub-)basins (Table 2). In the volcanic Almanor sub-basin, the largest shift corresponded to the 1987-92 drought (-25%), the longest in our record. For the surface-runoff-dominated East Branch, the largest shift was caused by the 1976-77 drought (-51%), the shortest, driest, and coolest in our record (see again Table 1); in general, this sub-basin consistently showed the largest shift during all droughts across all (sub-)basins. For the Feather River at Oroville, the largest ~~shift~~ shifts corresponded to the recent ~~2012-15 drought (~~2012-16 drought and the 1987-92 one (-21% and -22%~~), the warmest in our record (~~, respectively, note that no data was available for this basin during the 1976-77 drought). The 2007-09 drought showed the smallest shift in all basins under study.

3.3 PRMS performance for ~~flow~~streamflow

Both NSE$_{yrt}$ and LogNSE$_{yrt}$ significantly decreased during droughts (Figure 5). The difference between average NSE$_{yrt}$ and LogNSE$_{yrt}$ during drought vs. non-drought years and the four (sub-)basins was ~~-0.6 and -0.4~~ -0.55 and -0.35 (Almanor); ~~-1 and -0.4~~ -0.99 and -0.39 (East Branch); ~~-0.9 and -0.2~~ -0.88 and -0.21 (Middle Fork); and ~~-0.4 and -0.1~~ -0.35 and -0.11 (Feather at Oroville). The performance during isolated dry years was better than during droughts (e.g., see water years 1994 or 2001 in Figure 5). This decline in PRMS performance was comparable between droughts that were part of the calibration period (1970-1997) and those that occurred after 1997.

For the 1976-77 drought, LogNSE$_{yrt}$ $<$ NSE$_{yrt}$ in both the volcanic Almanor sub-basin and the East Branch. NSE$_{yrt}$ is very sensitive to high flows, while LogNSE$_{yrt}$ puts more weight on low flows (**?**): low flows were thus the driver of performance drops during the 1976-77 drought. In the East Branch, NSE$_{yrt}$ during that drought was even comparable to non-drought years. For the 1987-92 and the ~~2012-15~~ 2012-16 droughts, NSE$_{yrt}$ $<$ LogNSE$_{yrt}$ in both the Almanor sub-basin and the East Branch: high-flow peaks such as those during ~~winter precipitation events~~ intense, winter rain-on-snow events (frequent in these (sub-)basins, see **?**) or spring snowmelt were thus the main performance driver during these wetter droughts. The performance during the 2007-09 drought was only slightly below non-drought-year standards. The response time of NSE$_{yrt}$ to droughts in the Almanor sub-basin was somewhat slower than in the other (sub-)basins, a good example being the decadal drop during the 1980s ~~–~~ to early 1990s.

Observed annual full-natural flow was generally overestimated during all droughts but the 1976-77 one, for which PRMS severely underestimated water supply for both the Almanor and the East Branch sub-basins (relative biases of -0.44 and -0.86, respectively, Figure 6). Inter-annual variability in relative bias was larger in the surface-runoff-dominated East Branch and Middle Fork than in the volcanic Almanor sub-basin. The 2007-09 drought returned relative biases in the East Branch and Middle Fork that were in line with non-drought years.

PRMS overestimated the absolute value of climate elasticity of streamflow to both annual precipitation and potential evapotranspiration (Table 3 and Figure ~~S1~~ S3 in the Supporting Information). In particular, both observations and simulations showed a statistically significant positive elasticity to precipitation, but observations were closer to the theoretical elasticity according to the Turc-Mezentsev formula. With regard to annual potential evapotranspiration, observations did not show any statistically significant elasticity, whereas PRMS-based elasticity was statistically significant for both the Almanor and the East Branch sub-basins. ~~The largest overestimation of elasticity corresponded to the volcanic Almanor sub-basin.~~ We interpret the fact that modeled $e_{Q/PET}$ was unexpectedly positive as likely spurious and related to the large scatter and weaker correlations between potential evapotranspiration and modeled full-natural flow compared to precipitation vs. modeled full-natural flow (Figure ~~S1~~S3, univariate correlation coefficient for precipitation and potential evapotranspiration vs. modeled full-natural flow being $\sim$ ~~0.95~~ 0.94 and $\sim$~~0.43~~-0.43, respectively).

3.4 ~~The observed~~ Observed vs. modeled water balance

PRMS-modeled and PRISM-based basin-wide precipitation were visually in very good agreement, both in terms of annual values and in terms of inter-annual variability (Figure 7 and ~~S2~~ S4 - ~~S3~~ S5 in the Supporting Information). This outcome was expected, as PRMS uses PRISM as a starting point to distribute precipitation across the watershed (DRAPER method, see Section 2.3.2), which does not significantly affect annual precipitation totals. On the other hand, the model significantly underestimated annual estimated evapotranspiration, even if this underestimation was partially compensated for by the groundwater sink (Figure 7 and ~~S2~~ S4 - ~~S3~~S5). Also, the model systematically underestimated ~~both the absolute value and the interannual variability of changes~~ the interannual variability in sub-surface storage~~, in particular for~~ ; in the Almanor sub-basin ~~(Figure S3) and for the Feather River at Oroville(Figure 7); PRMS~~ and the Feather at Oroville, PRMS also failed to reproduce the estimated multi-decadal decline in storage~~observed in all (sub-)basins as a result. While the observed~~. While estimated changes in sub-surface storage ~~used in this paper to evaluate PRMS~~ may suffer from unquantifiable uncertainty across precipitation, full-natural flow, and ~~GAM-estimated~~ evapotranspiration, this decline was confirmed by other ~~soft data~~ pieces of evidence collected on the river ~~(?)~~ (such as a decline in spring outflow, see **?**), and agrees with a general trend of declining summer low flows in the Maritime Western U.S. (**?**).

Among the three water-balance components determining full-natural flow, ET was the only one for which the performance of PRMS during drought and during non-drought years were statistically different in all (sub-)basins (see Figure 8 and Table 4). Conversely, the performance for precipitation and for changes in sub-surface storage were statistically different in only one sub-basin each: the East Branch for P and Almanor for ΔS, respectively. As expected (see Section 3.3), differences in PRMS performance for full-natural flow during drought vs. non-drought years were statistically significant in all basins. Thus, ET was the only water-balance component (besided Q) that was systematically misrepresented during droughts.

The statistically different performance of PRMS for ET during droughts was confirmed even when comparing simulated and observed average ET (including groundwater sink) over temporal windows of two, three, and four years (see Figure ~~S4~~ S6 in the Supporting Information), which may be more appropriate time scales for this evaluation because of the possible temporal lag between vegetation greenness and ET (**?**). Results in Figure 8 and Table 4 were also confirmed when only considering ET rather than the sum of ET and the groundwater sink (see Figure ~~S5~~ S7 in the Supporting Information and Section 2.3.3 for details about the groundwater sink).

4 Discussion

~~Previous studies about drought-driven shifts in precipitation vs. runoff have mostly focused on rainfall-dominated regions like Australia (see **?**, and references therein) or China (**?**), but we showed here that such shifts may also occur in mixed rain-snow catchments in a Mediterranean climate, regardless of predominant geology (volcanic, subsurface-flow-dominated or transitional-to-granitic, surface-runoff-dominated). In agreement with previous findings by **?**, we also found that only droughts corresponding to significant shifts in~~

Multi-year droughts can trigger shifts in the precipitation-runoff ~~relationships translated into poorer performances for a semi-distributed hydrologic model (PRMS), meaning that understanding this drop in accuracy may shed light on the mechanism behind the observed shifts. Because~~ relationship, which is the fundamental allocation rule of a basin's water budget (**????**). Major droughts reduce P, which directly reduces Q through non-linear feedback mechanisms involving ET and ΔS. By showing that ET is the only water-balance component ~~(besides Q)~~ yielding statistically different performances during droughts vs. non-drought years ~~,~~ for a hydrologic model (PRMS), we quantitatively demonstrated that ET-drought feedback mechanisms are the ~~most~~ likely driver of these shifts in water supply~~in a Mediterranean, mixed rain-snow climate, and failure~~. Failure to fully account for these non-linear mechanisms results in the predictive inaccuracy of the model ~~.~~ during droughts. We focused on mixed rain-snow catchments with asynchronous precipitation and growing seasons, thus extending the climatic range of previous drought studies that identified shifts in precipitation-runoff relationships of comparable magnitude in rainfall-dominated semi-arid areas (**??**, and references therein).

~~Shifts in precipitation vs. runoff have attracted increasing interest since at least the recent Millennium drought in south-eastern Australia, where **?** estimated maximum shifts in precipitation vs. runoff in the order of 80-100%.~~ If and where ~~these shifts~~ shifts in the precipitation-runoff relationship occur, runoff will decrease more than what would be predicted based on a precipitation-runoff relationship trained using non-drought years (**?**). With hydrologic models that are often biased toward better reproducing high flows (**?**) and climate-change scenarios that predict increasing aridity in several regions of the world (**?**), understanding the mechanisms behind these shifts and the ~~adequacy (**?**) of models in~~ predictive skill of models under such conditions has profound implications for water resources management and water security.

~~In this Section~~In this section, we elaborate on three of these implications: first, ~~why is ET consistently misrepresented during droughts? Second,~~ are shifts in precipitation vs. runoff common across all basins of the Sierra Nevada? Second, how do ET-drought feedback mechanisms cause shifts in the water balance? Third, what are the lessons learned to improve water-supply simulations in drought-prone regions?

~~Annual errors for basin-wide ET were due to (1) a systematic bias ($\sim$160 mm less simulated ET than observed) and (2) an annually variable component (Figure 10, results refer to the Feather at Oroville). The systematic bias could be explained by an underestimation of plant-accessible water storage, a recurring source of uncertainty in the~~ Are shifts in precipitation vs. runoff common across the Sierra Nevada~~(**?**). Like any hydrologic model that is calibrated on streamflow, however, the annual performance of PRMS for Q is relatively insensitive to systematic biases in internal fluxes like ET as these can be offset by other fluxes during calibration (a good example being the groundwater sink, see Figure 7). This means that the drop in modeling~~

~~accuracy during droughts and its relationship with shifts in precipitation-runoff relationships is related to the annually variable component. This component of the error was indeed qualitatively correlated with drought vs. non-drought years: PRMS tended to relatively overestimate ET during droughts and to underestimate it during non-drought years (see again Figure 10). We did not find any qualitative correlation between errors for ΔS and drought years (Figure S6), implying again that ET is the driver of predictive inaccuracy for this model during droughts.~~ ?]~~Why ET is misrepresented during droughts: climate elasticity of evapotranspiration~~

~~Annual errors for basin-wide ET were due to (1) a systematic bias ($\sim$160 mm less simulated ET than observed) and (2) an annually variable component (Figure 10, results refer to the Feather at Oroville). The systematic bias could be explained by an underestimation of plant-accessible water storage, a recurring source of uncertainty in the~~ Are shifts in precipitation vs. runoff common across the Sierra Nevada~~(**?**). Like any hydrologic model that is calibrated on streamflow, however, the annual performance of PRMS for Q is relatively insensitive to systematic biases in internal fluxes like ET as these can be offset by other fluxes during calibration (a good example being the groundwater sink, see Figure 7). This means that the drop in modeling accuracy during droughts and its relationship with shifts in precipitation-runoff relationships is related to the annually variable component. This component of the error was indeed qualitatively correlated with drought vs. non-drought years: PRMS tended to relatively overestimate ET during droughts and to underestimate it during non-drought years (see again Figure 10). We did not find any qualitative correlation between errors for ΔS and drought years (Figure S6), implying again that ET is the driver of predictive inaccuracy for this model during droughts.~~ ?

~~**?** suggested three sources of conceptual uncertainty with regard to how models simulate the drought water balance. The first is the already mentioned oversimplification of regolith storage and rooting depth and thus misrepresentation of plant-accessible water storage (see also **?**). The second is a lack of proper parametrization of the feedback between evaporative demand and stomatal closure. The third is the representation of vegetation as a static layer with no dynamic response to stresses. From this perspective, Figure 5 shows that relative differences between observed and simulated ET during droughts across the Feather at Oroville (relative to the systematic bias) were somewhat correlated with maximum and minimum annual temperature (r = -0.45 and -0.57, respectively) and with observed relative changes in storage ($r = 0.23$ – also relative to the corresponding systematic bias, see Figure S6). The correlation with annual precipitation during droughts was much smaller ($r = 0.1$). Correlations during non-drought years were visually similar to those during droughts (Figure 5), and differences compared to drought years should be interpreted with caution given the small number of available data points.~~

~~While none of these correlations is strong enough alone to explain modeling errors, they collectively point to interactions across atmospheric demand for moisture, ET, and sub-surface storage as the source of conceptual uncertainty behind the misrepresentation of ET during droughts. The overall picture is that PRMS relatively overestimates ET during years characterized by comparatively cold conditions and a relative replenishment of sub-surface storage (both conditions that would in fact decrease ET), and underestimates ET during comparatively warm years characterized by a relative drawdown of storage (both conditions that would in fact increase ET). In other words, the model underestimates the multi-year response of ET to climate *variability*, a property that we hereby define as *climate elasticity of evapotranspiration* and that emerges as a driver of~~

~~water supply in a Mediterranean climate (?). While tree mortality may also be a good explanatory variable for errors in ET, available data on this river only date back to the early 2000s and were thus too short to compute correlations.~~

4.1 ~~Are shifts in precipitation vs. runoff common across the Sierra Nevada?~~

~~Our results show that shifts in~~ Sierra Nevada geology transitions from crystalline igneous intrusive rocks (granitic) in the southern and central Sierra Nevada to more porous igneous extrusive rock in the northern Sierra Nevada. In the Feather River, northern Sierra Nevada, the Almanor sub-basin is comprised of extrusive igneous bedrock and is subsurface-flow dominated, whereas the East Branch sub-basin is transitional to granitic rock and is surface-runoff dominated. Shifts in the precipitation-runoff relationship ~~may take place both in volcanic, subsurface-flow-dominated and in transitional-to-granitic, surface-runoff-dominated basins~~in response to droughts occurs in both sub-basins. This may seem counterintuitive as the ~~volcanic~~ Almanor sub-basin has a relatively small inter-annual variability in low flows~~that agrees with other studies in similar contexts (???), and this variability is an important driver of shifts in Australian basins (?). However, while the surface-runoff-dominated East Branch does return larger shifts both overall (M_Q) and for individual droughts (m_Q, see Section 3.2), both this sub-basin and the Almanor~~, similar to other subsurface-flow dominated basins (**???**). Both the Almanor and the East Branch sub-basins, however, are rain-shadowed (aridity index of ∼1.5 and 1.1, respectively). ~~According to ?, pre-drought~~ Pre-drought aridity is the most important predictor of shifts according to **?**. This highlights that a higher summer flow does not necessarily provide resiliency to shifts in precipitation vs. runoff during droughts. ~~That said, the volcanic~~ Still, the Almanor sub-basin ~~was still~~ has been more sensitive to longer droughts than the ~~surface-runoff-dominated~~ East Branch sub-basin. This behavior may be related to the comparatively long time needed by groundwater-flow-dominated, slow-draining catchments to respond to a superficial precipitation deficit ~~, which has also been shown to decrease elasticity of summer low flows to superficial inputs (?)~~(**?**).

~~The magnitude of shifts in our case study were comparable to previous findings: ? reported an average shift of -24% in 18 rivers in China; ? found a mode between -40% and -20% in Australia. By upscaling the analysis in the present study to the twelve~~ To test the sensitivity of western slope Sierra Nevada rivers to shifts in precipitation-runoff relationships in response to droughts, we extended our analysis to the 12 other major rivers~~draining the western side of the Sierra Nevada to the California Central Valley, we found that eight~~. We found that nine of these twelve basins showed statistically significant shifts, on the order of ~~-19% to -12~~-20% to -10% (Figure 9 and Table ~~S3~~S4, water years 1985-2018, data from PRISM and https://cdec.water.ca.gov/, visited July 19, 2019). The magnitude of these shifts are comparable to previous studies, -24% on average in 18 rivers in China (**?**) and a mode between -40% and -20% in Australia (**?**). The basins that did not show a statistically significant shift were the relatively small, low-elevation ~~Cosumnes and Tule basins plus~~ Tule basin and two other southern basins, the Kaweah and the Kern (see a map and a summary of hydrologic characteristics of each basin in **?**). While the Kaweah and the Kern have high-elevation, snow-dominated headwaters, they also have significant rain-modulated low-elevation areas. ~~Likewise, the low-elevation Cosumnes and Tule are mostly rain dominated.~~ This suggests that some mixed rain-snow, Mediterranean basins in which rain plays a ~~more~~ prominent role in the annual water budget are less prone to shifts in the precipitation-runoff relationship. ~~We interpret this as being because, in~~ In the snow-dominated basins where a significant shift is observed, snow-melt replenishes sub-surface storage later into the dry season and thus limits the dependence of evapotranspiration from deep sub-surface storage (**?**), a key mechanism that ~~also~~ reduces low-flow elasticity to climate variability (**?**) ~~but~~ and that is greatly reduced during longer droughts.

~~The significance of these shifts was, however, sensitive to a number of methodological choices that are worth discussing given the relatively small amount of literature on statistically quantifying~~

4.1 Climate elasticity of ET and hysteresis of the precipitation-runoff relationship

Annual PRMS errors for basin-wide ET were manifest in both a systematic bias ($\sim$110 mm less simulated ET than observed) and an annually variable component (Figure 7, results refer to the Feather at Oroville). The systematic bias could be explained by an underestimation of plant-accessible water storage, a recurring source of uncertainty in the Sierra Nevada (**?**). Like any hydrologic model that is calibrated on streamflow, the annual performance of PRMS for Q is relatively insensitive to systematic biases in internal fluxes like ET, as these can be offset by other fluxes during calibration (a good example being the groundwater sink). This means that the drop in modeling accuracy during droughts and its relationship with shifts in precipitation-runoff relationships ~~. First, if the period 2016-2018 had been removed from computations in Figure 9 as was done for the Feather in Figure 4, only the Feather,~~ is related to the second component of the error (the annually variable one). This second component of the error was visually well correlated with annual precipitation and, in particular, with wetting-drying cycles (Figure 10, top panel), with the model generally overestimating ET during wetting cycles and underestimating it during drying cycles. This suggests that the model consistently underestimated the buffered response of ET to precipitation variability, a property that we refer to as the *climate elasticity of evapotranspiration* and that emerges as a driver of water supply in a Mediterranean climate (**?**).

? identified three sources of conceptual uncertainty about how models simulate evapotranspiration. The first is the oversimplification of regolith storage and rooting depth and thus misrepresentation of plant-accessible water storage (see also **?**). The second is a lack of proper parametrization of the feedback between evaporative demand and stomatal closure. The third is the representation of vegetation as a static layer with no dynamic response to stresses. Across the Feather at Oroville, differences between estimated and simulated ET were well explained by estimated changes in storage ($r = 0.78$, see the bottom panel in Figure 10). The modeling missing piece behind the misrepresentation of climate elasticity of evapotranspiration and therefore behind the observed shifts in the water balance was thus the multi-year carryover of soil-water storage. This storage is a critical drought-resilience source for asynchronous Mediterranean catchments buffering ET temporal patterns from the ~~Tuolomne, and the San Joaquin would have returned a statistically significant shift. The magnitude of the shift was more robust, with average differences between the estimates with or without the period 2016-18 on the order of 3%. Second, annual precipitation in Figure 9 was directly estimated from PRISM surfaces rather than by tilting them with index in-situ stations as done on the Feather (see Section 3.3). This choice was made for consistency reasons due to the lack of a comparable tool to DRAPER in the other basins of the Sierra. While results for the Feather River at Oroville using both approaches were visually in good agreement (not shown), the absolute magnitude of the shift for this river is smaller with original (-12%)than with tilted PRISM data (-18%). It is challenging to assess which of the two indexes is closer to actual precipitation , but this comparison helps quantify the contribution of P on the overall uncertainty in M_Q. Third, results were sensitive to the choice of using the~~

~~Box-Cox transformation (Equation 2): by focusing on the Feather river (~~lack of precipitation through the accumulation of plant-accessible water during wet periods and drawdown during dry periods (**??**). This buffering process was not reproduced by the model, which consistently underestimated inter-annual variability in ΔS for all (sub-~~)basins, only the East Branch and the Feather at Oroville would have returned a statistically significant shift using non-transformed full-natural flow data.)~~basins (Figure 7) and overestimated streamflow elasticity to precipitation (Table 3) as a result.

~~Overall, this discussion demonstrated that precipitation vs. runoff shifts are a common feature of mixed~~ Interactions between ΔS and ET thus emerge at the ultimate cause of shifts in the water balance of mixed rain-snow~~catchments in the Sierra Nevada of California. At the same time, results underscored the importance of explicitly including an uncertainty-sensitivity analysis when quantifying these shifts, especially because they are the result of differences across large numbers (~~, Mediterranean regions during droughts. Reinterpreting shifts as evidence of *hysteresis* in the precipitation-runoff relationship yields a process occurring in four distinct phases (Figure 5). During the initial stages of a drought, multi-year carryover from previous wet periods can compensate for missing precipitation input, a mechanism that offsets ET deficit and therefore maintains proportionality between precipitation and runoff (phase 1, see Figure 5). This offsetting mechanism is predominant during an isolated dry period, while during a prolonged drought drawdown leads to soil-water depletion and an associated drop in runoff due to preferential allocation of surface water to ET (**?**); this ultimately leads to catchment coevolution processes such as tree mortality (**??**) – phase 2. With the loss of buffering capacity, water basins will respond to precipitation according to a different precipitation-runoff relationship (phase 3), because of a different allocation of P ~~and~~ across ET, Q ~~) that are particularly uncertain in mountain contexts (?)~~and ΔS. Recovery during following wet periods will eventually lead to other catchment coevolution processes such as vegetation expansion, which will possibly restore the initial precipitation-runoff relationship (phase 4, end of hysteresis).

These four phases associated with precipitation-runoff hysteresis have never been investigated in a systematic way, but recent Critical-Zone studies in the Kings river, California show that they were the predominant mechanisms controlling water-supply changes during droughts (**?**). We believe them to apply to the higher-precipitation, transition geology of the Feather river basin and across the Sierra Nevada. Recent pan-European studies have also showed persistence of ET for longer temporal scales than runoff during droughts (**?**), while modeling studies across the European Alps have showed that ET can even increase during droughts if air temperature increases (**?**). Better understanding hysteresis in the precipitation-runoff relationship is, therefore, a priority of future work, especially in a warming climate.

4.2 How to achieve more robust runoff predictions during droughts?

~~While we considered only one model, PRMS, the conceptual-uncertainty~~ The conceptual uncertainty source outlined in Figure 5 is a common feature among hydrologic models that traces back to fundamental knowledge gaps in ~~Critical Zone science such as the actual depth to which roots can access water in regolith (?). While it has been hypothesized~~ critical-zone science (**?**), including their tendency to systematically underestimate multi-year carryover of soil-water storage and therefore miscapture interannual variability in subsurface flow (see Figure 7 and **?**). Thus, predictions of hydrologic models during droughts should be interpreted with caution.

While **?** hypothesize that more observations could improve the performance of models during droughts~~(**?**),~~ , we found that the accuracy of PRMS for droughts during the calibration period was similar to that for droughts outside it. ~~This version of PRMS was calibrated by mixing visual inspection and multiple objective functions such as root mean square error, bias, and relative error (**?**), which may have skewed model fitting toward high flows compared to multi-objective criteria like the Kling-Gupta Efficiency (**?**) or low-flow-oriented metrics like LogNSE.~~ More measurements of ~~evapotranspiration~~ the water balance in mountain regions could help better parametrize climate elasticity of evapotranspiration and ~~thus support improved calibration by~~ multi-year carryover of soil-water storage, thus unraveling the role that ~~this property~~ these mechanisms plays in buffering the impact of precipitation deficit on runoff during droughts ~~.~~ and supporting improved process representations.

Shifts in the precipitation-runoff relationship of snow-dominated regions are ~~particularly~~ critical because in these contexts snow plays a key role in both water supply and its seasonal predictability. In the western United States, water-supply forecasts are based on statistical regressions across full-natural flow, precipitation, and snowpack accumulation (**???**). These forecasts play a key role in water-resources allocation across industrial and agricultural uses as well as freshwater supply (**?**). While using runoff-to-date as a predictor and fitting different regression coefficients for different year types may partially correct for runoff deficit, these regressions do not explicitly account for shifts in precipitation-runoff relationships during droughts (**?**). Since shifts in precipitation-runoff relationships are common across the Sierra Nevada (Figure 9), we suggest embedding a shift predictor into these regressions as done in Equation 1 as ~~a potential future step of this work~~the next step to improve water-supply forecasts.

The underestimation of runoff during the 1976-77 drought disagrees with the general consensus that these models tend to overestimate water supply in regions where droughts shift the precipitation-runoff relationship (**??**). Water year 1977 was still the driest among the 114 years on record for California in 2015 (**?**), and it was the last of three years of consecutively decreasing precipitation (Figure 2). Long dry periods may lead to a disconnection between soil and groundwater storage, which in turn may prevent recharge and favor direct surface runoff and interflow (see **?**, and references therein). This condition ~~of temporary hydrophobicity of soils~~ and the subsequent slower-than-expected recovery of soil-water storage (**?**) are not captured by PRMS. Neglecting this process may lead to erroneously allocating precipitation to the recharge zone (where it becomes available for evapotranspiration) or to groundwater; in either case, runoff would be underestimated. Here again, surface-to-subsurface mass and energy fluxes emerge as the most relevant knowledge gap in this field that would benefit from more targeted research.

5 Conclusions

Droughts cause a shift in the precipitation-runoff relationship of Mediterranean mixed rain-snow mountain basins of the Sierra Nevada of California. The magnitude of this shift is comparable to previous findings in rainfall-dominated semi-arid areas with year-around or summer-monsoon-dominated precipitation, which points to common feedbacks impacting the process across precipitation regimes, regardless of geology. By comparing observed water-balance components during drought vs. non-drought years with those simulated by a hydrologic model, we identified ~~some of~~ these common feedbacks as being driven by the ~~multi-year~~ interplay between the response of evapotranspiration to climate ~~and in particular to atmospheric~~

~~demand for moisture (temperature) and to~~ (climate elasticity of evapotranspiration) and multi-year carry-over of subsurface water storage. ~~Surface-runoff-dominated catchments are prone to larger shifts in precipitation-runoff relationships than catchments dominated by subsurface flow because of the modulating effect of groundwater on the annual water balance of the latter. Snow-dominated basins are also more susceptible to shiftsthan rainfall-dominated basins because snow melt during the dry season limits evapotranspiration dependence on deep sub-surface storage – a mechanism that is greatly reduced during droughts.~~ Shifts therefore are a hysteretic response of the water budget to buffered catchment-climate coevolution mechanisms like tree mortality and expansion and soil-water drawdown and replenishment. These processes take place over comparatively long, multi-year time scales, thus explaining why short dry periods are often not subjected to shifts. The complex response of evapotranspiration to climate in mixed rain-snow Mediterranean catchments caused significant drops in performance for a ~~semi-distributed~~ hydrologic model (PRMS). Improved parametrizations of climate elasticity of evapotranspiration are thus highly needed to make models and water resources management more robust to droughts, especially in a warming and more variable climate. A primary need in this regard is to better represent the buffering role of deep, plant-accessible subsurface water storage during multi-year dry periods in sustaining vegetation evaporative demand.

Competing interests. Authors have no competing interest.

Code and data availability. Sources of data used in this paper are reported in the main text (in particular in Section 2.2). The PRMS hydrologic model is available at https://www.usgs.gov/software/precipitation-runoff-modeling-system-prms.

Acknowledgements. This work was supported by the California Energy Commission under contract EPC-14-067, Pacific Gas & Electric Co, the California Department of Water Resources, ~~and~~ the UC Water Security and Sustainability Research Initiative, and the National Science Foundation through the Southern Sierra Critical Zone Observatory (EAR-1331939) and a Major Research Instrumentation grant (EAR-1126887). Tessa Maurer was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1106400. We would like to thank Kevin Richards (PG&E) for his continuous support and advice throughout this work.

Table 1. Average precipitation (P), Snow Water Equivalent (SWE), and maximum-minimum daily temperature (T_{max} and T_{min}, respectively) across the four California droughts under study.~~Annual statistics are reported in Figure 2.~~

Drought	P (m)	SWE (m)	~~SWE/P (-)~~T_{max} (degC)	T_{min} (degC)[a]
1976-77	0.49	0.23	~~0.46~~ 19.56	2.66
1987-92	~~0.79~~ 0.77	0.39	~~0.51~~ 20.08	3.55
2007-09	0.90	0.46	~~0.51~~ 19.30	4.02
~~2012-15~~ 2012-16	~~0.84~~ 0.95	~~0.22~~ 0.29	~~0.26~~ 19.93	~~20.04 4.58~~4.64

---

[a] P is average water-year precipitation during each drought across 13 stations on the Feather River (see Table ~~S1~~ S2 in the Supporting Information). SWE is average April 1st SWE during each drought across 25 stations on the Feather River (see Table ~~S2~~ S3 in the Supporting Information). T_{max} and T_{min} are average annual maximum and minimum daily temperature during the drought at the three index stations used by the PRMS model for air-temperature input: Canyon Dam (1390 m), Quincy (1066 m), and Bucks Creek Powerhouse (536 m). Data sources: California Data Exchange Center (CDEC, https://cdec.water.ca.gov/, visited July 19, 2019) and Pacific Gas & Electric.

Table 2. Estimated shift in the precipitation-runoff relationship for single droughts on the Feather River (see Section 2.3.1).

Drought	m_Q Almanor (%)	m_Q East Branch (%)	m_Q Oroville (%)
1976-77	-11	-51	–
1987-92	-25	-36	~~-21~~ -22
2007-09	-9	-20	-6
~~2012-15~~ 2012-16	-18	~~-47~~ -44	~~-22~~ -21

Table 3. Modeled, observed, and theoretical climate elasticity of streamflow to annual precipitation ($e_{Q/P}$) and potential evapotranspiration ($e_{Q/PET}$) for the three (sub-)basins under study with complete annual full-natural flow data. Theoretical elasticity was computed according to the Turc-Mezentsev formula (**?**). The asterisk (*) denotes statistically significant elasticity values (that is, the sign of the confidence bounds agrees, 95% confidence level).

(Sub-)basin	$e_{Q/P}$ (-)			$e_{Q/PET}$ (-)		
	Modeled	Observed	Theoretical	Modeled	Observed	Theoretical
Almanor	~~0.80~~0.79*	~~0.59~~0.58*	0.55	~~0.44~~0.36*	~~-0.08~~ -0.17	-0.31
East Branch	0.68*	0.56*	0.33	~~0.21~~0.24*	~~-0.08~~ 0.00	-0.14
Oroville	0.73*	0.69*	~~0.47~~ 0.48	~~0.26~~ 0.27	~~-0.07~~ -0.06	-0.25

Table 4. Regression between observed and simulated water-balance components: confidence bounds of the regression parameter ruling shifts in performance during droughts (see Equation 4). ET is the sum of ET and the groundwater sink (see Section 2.3.3). The asterisk (*) denotes statistically significant elasticity values (that is, the sign of the confidence bounds agrees, 95% confidence level).

(Sub-)basin	$b_{1,P}$ (mm)	$b_{1,ET}$ (mm)	$b_{1,\Delta S}$ (mm)	$b_{1,Q}$ (mm)
Almanor	~~-136~~ -119 to +~~42~~ 51	~~-125 to -33~~ -102 to -20*	+~~13~~ 35 to +~~119~~ 155*	-154 to ~~-1~~ -9*
East Branch	~~-187 to -14~~ -191 to -30*	~~-81 to -17~~ -80 to -14*	~~-57~~ -83 to +~~41~~ 47	~~-129 to -17~~ -128 to -24*
Oroville	~~-125~~ -96 to +~~4~~ 56	~~-76 to -14~~ -80 to -9*	~~-5~~ -19 to +~~76~~ 142	~~-117 to -4~~ -120 to -14*

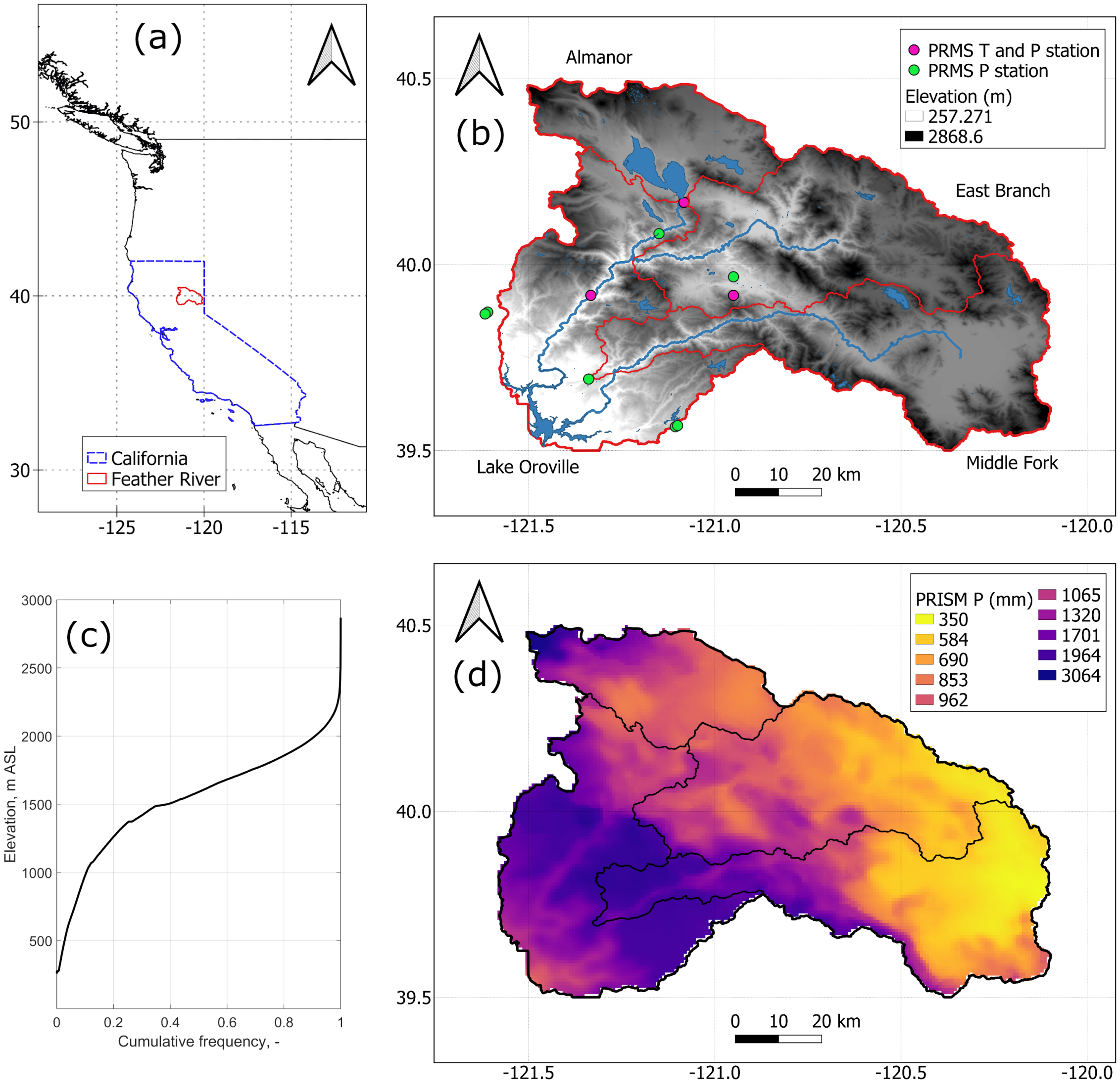


Figure 1. The Feather River at Oroville and its three headwater sub-basins under study (Almanor, East Branch, and Middle Fork): location of the river along the Pacific coast of North America (a), orography and hydrography, along with in-situ temperature and precipitation stations used to force PRMS (~~a~~see Section 2.3.2) – (b), cumulative-frequency distribution of river-basin's elevation (c), and PRISM 1981-2010 average annual precipitation (~~b~~d)~~, predominant geology according to the USGS National Atlas (c)~~.

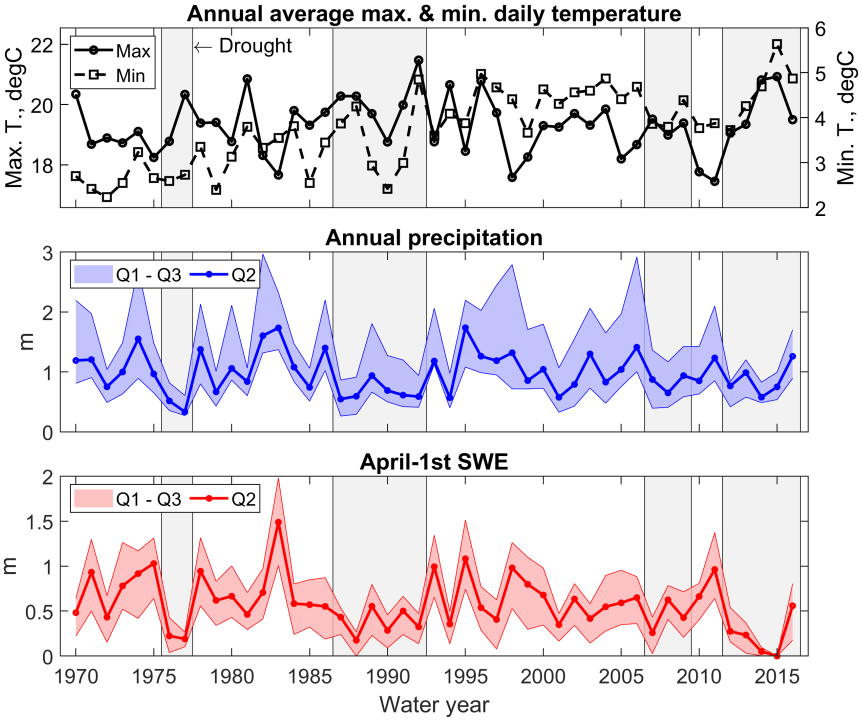


Figure 2. Hydroclimatic summary of the four most recent Californian droughts compared to non-drought years. Maximum and minimum daily temperature is an annual average of the three index stations used by the PRMS model for air-temperature inputs: Canyon Dam (1390 m), Quincy (1066 m), and Bucks Creek Powerhouse (536 m). Annual precipitation and April-1st SWE were computed using 13 and 25 stations across the Feather River, respectively (see Table ~~S1 and~~ S2 and S3 in the Supporting Information). ~~The ratio between April-1st SWE and annual precipitation was computed with reference to spatial medians.~~ Q2 is the spatial median, Q1 and Q3 are the two quartiles, respectively. Sources: California Data Exchange Center (CDEC, https://cdec.water.ca.gov/, visited July 19, 2019) and Pacific Gas & Electric.

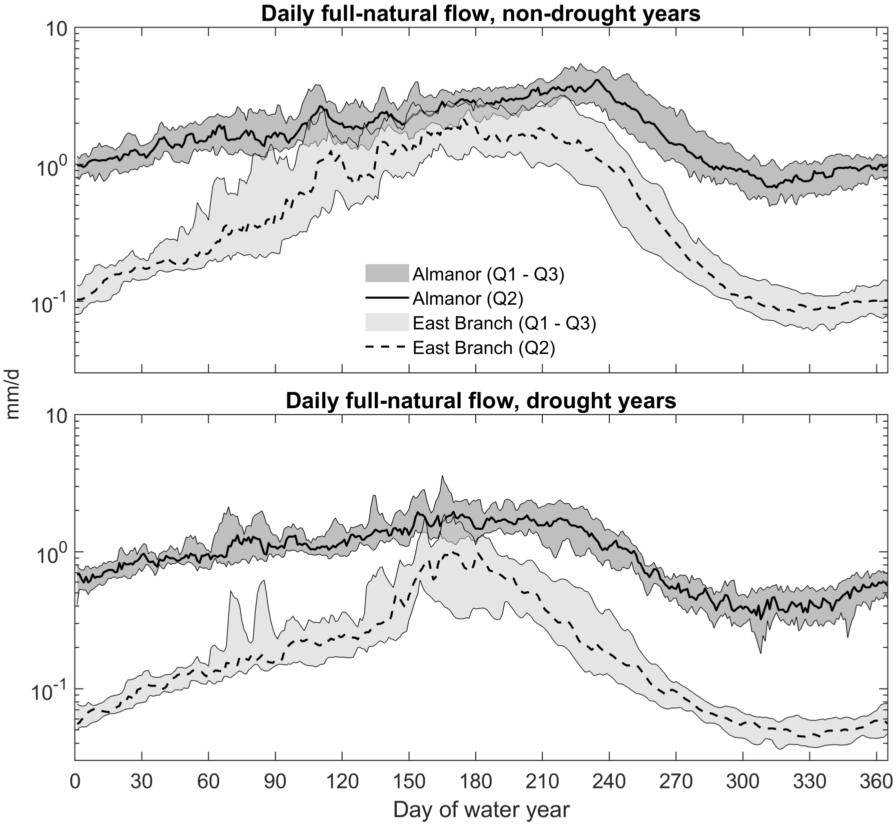


Figure 3. Daily median (Q2) and quartiles (Q1 and Q3) of observed full-natural flow during drought and non-drought years at the outlet of two headwater basins of the Feather River with contrasting geology. The Almanor subbasin is a predominantly volcanic, subsurface-flow-fed subbasin; the East Branch is transitional to granitic and surface-runoff-dominated. The y axis in in log-scale.

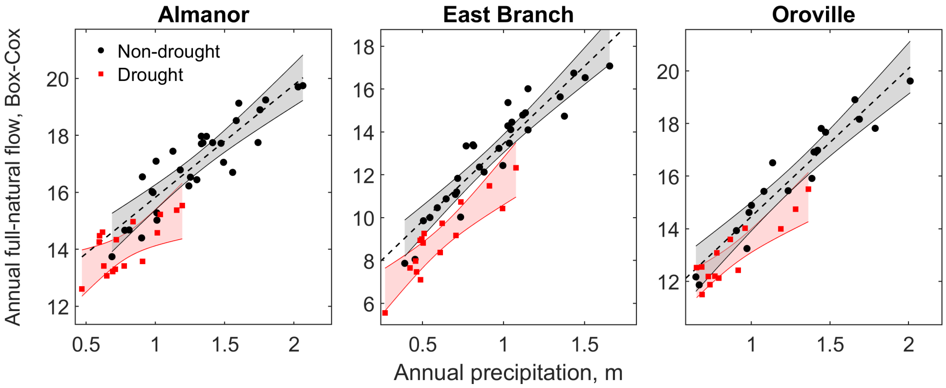


Figure 4. Precipitation-runoff relationship for drought (red) vs. non-drought (black) years and the three basins under study with complete annual runoff data (Almanor, East Branch, and Feather River at Oroville). The red and grey bands represent 95% confidence intervals for the regressions. The Box-Cox transformation for annual full-natural flow is introduced in Section 2.3.1, Equation 2.

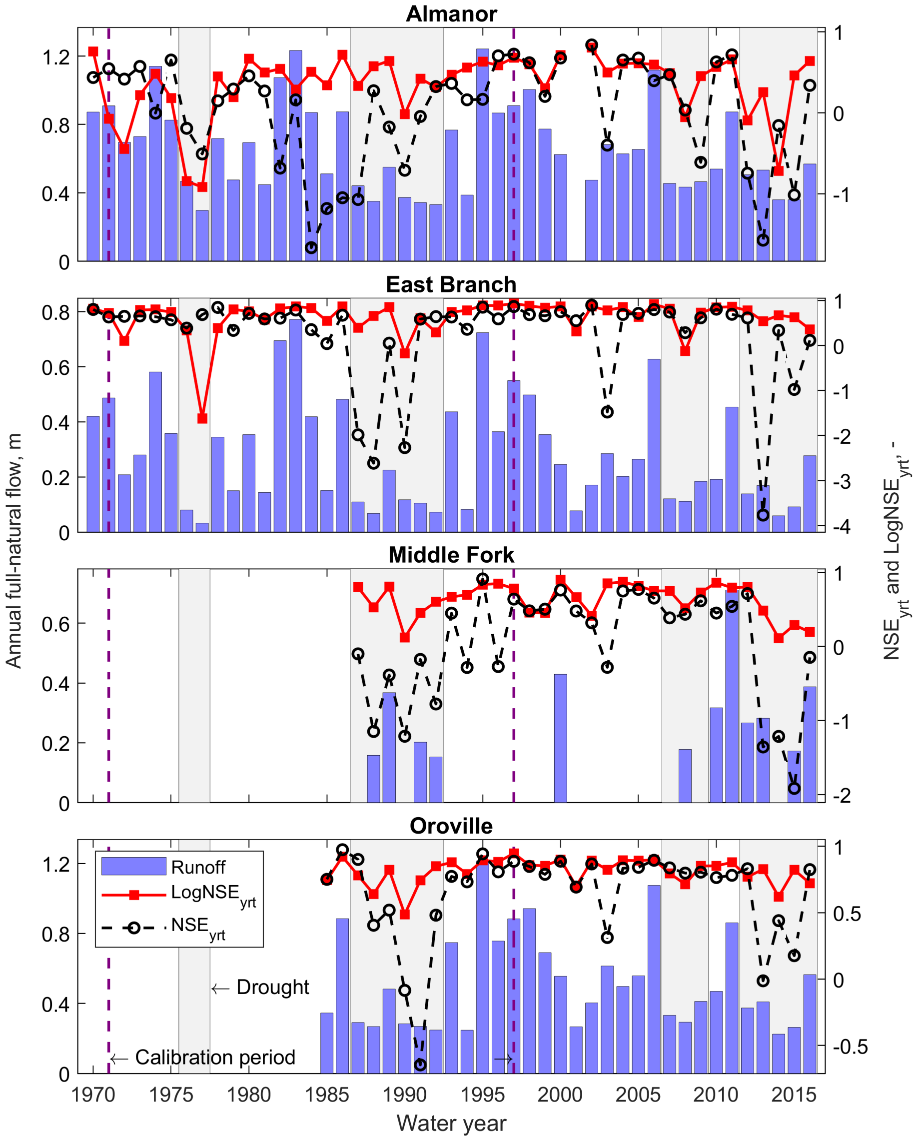


Figure 5. Water-year Nash-Sutcliffe Efficiency (NSE_{yrt}) and Log-Nash-Sutcliffe Efficiency ($LogNSE_{yrt}$) for PRMS-modeled daily full-natural flow. The blue bars represent observed annual full-natural flow at the outlet of each (sub-)basin.

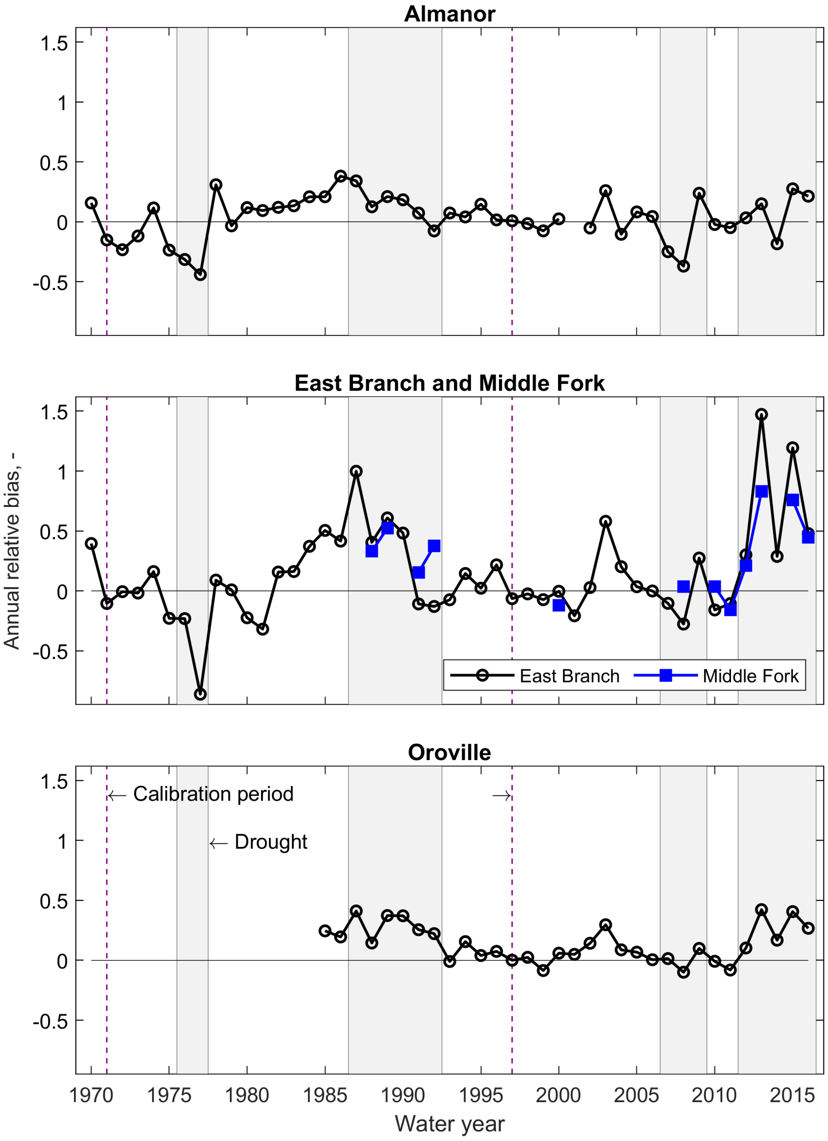


Figure 6. Annual relative bias for full-natural flow at the outlet of all (sub-)basins under study (difference between model and observations of cumulative annual full-natural flow, relative to the latter).

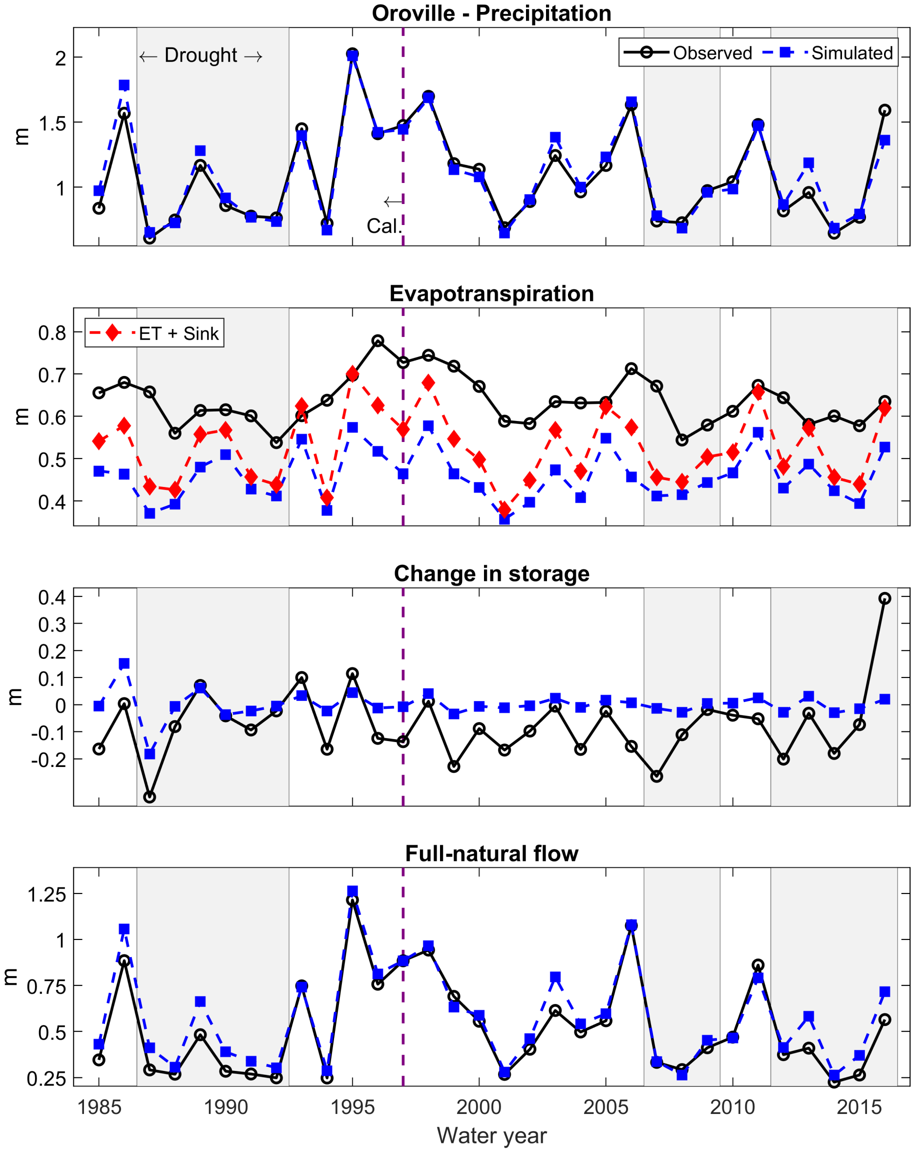


Figure 7. Simulated vs. observed annual basin-wide water-balance components (P, ET, ΔS, and Q) for the Feather River at Oroville.

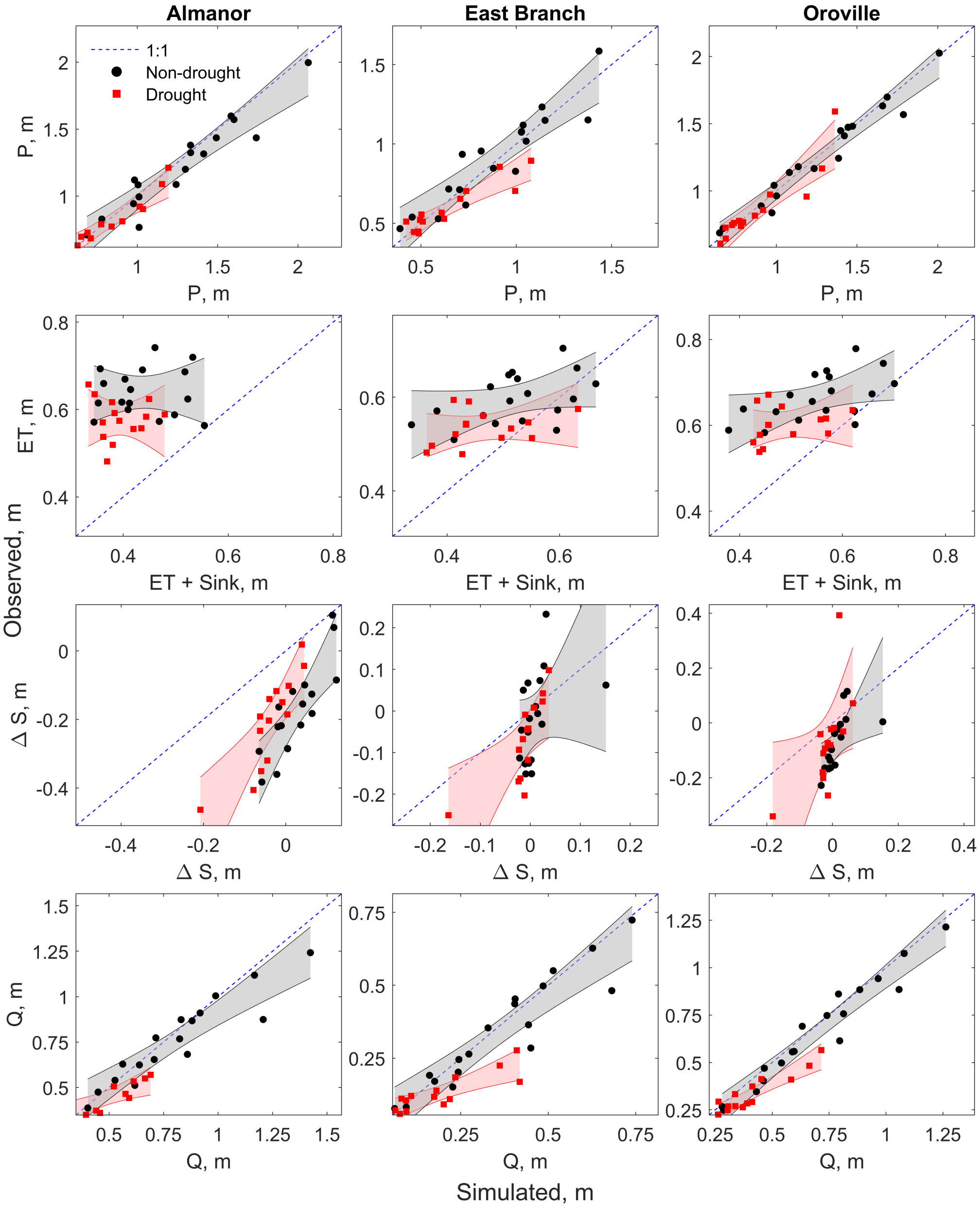


Figure 8. ~~Scatter plot of simulated~~ Simulated vs. observed annual basin-wide water-balance components (P, ET, ΔS, and Q) separated between drought (red) and non-drought (black) years. Simulated annual ET includes the groundwater-sink mass-flux component (see Section 2.3.3). The red and grey bands represent 95% confidence intervals for the regressions.

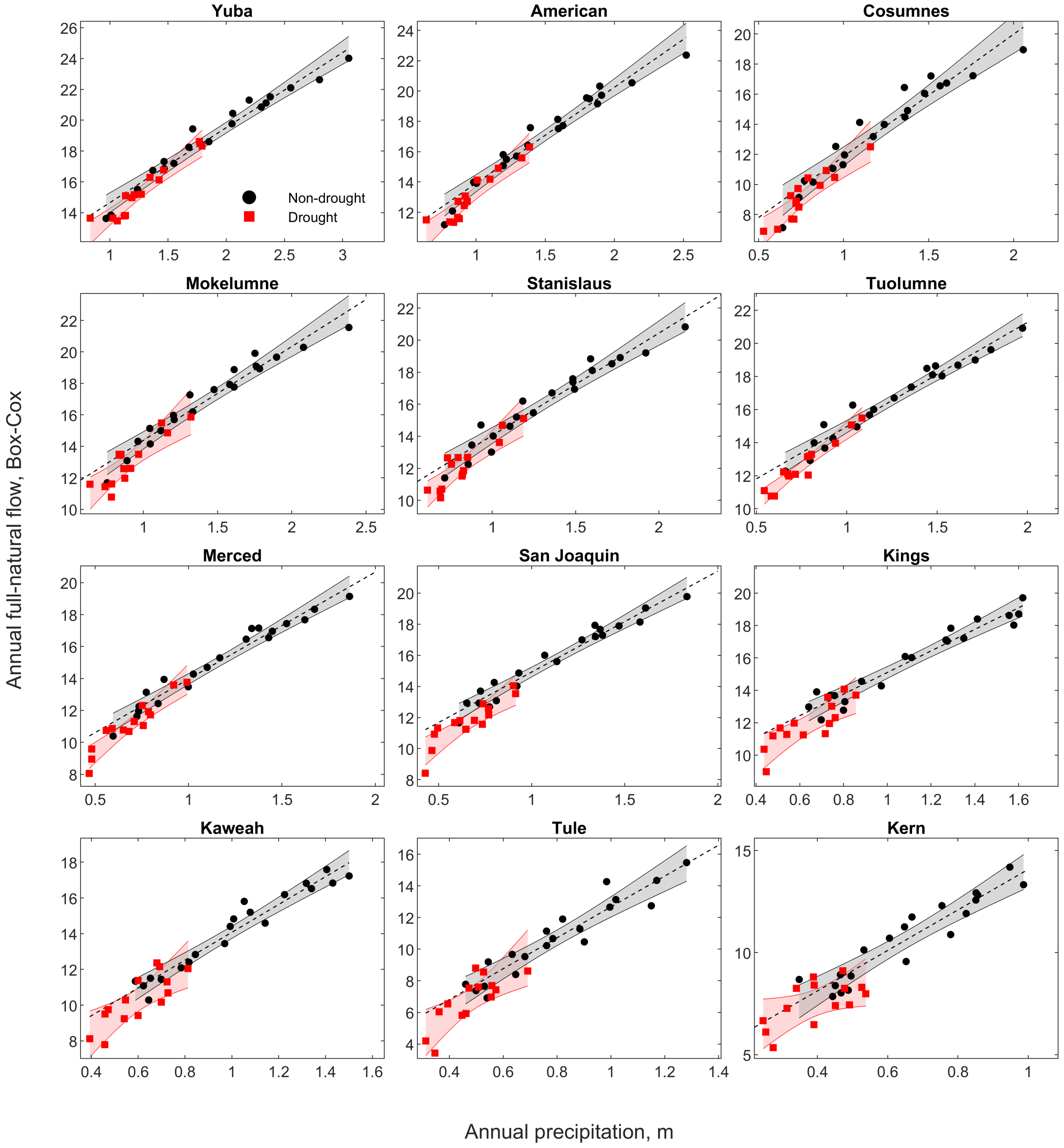


Figure 9. Precipitation-runoff relationship for drought (red) vs. non-drought (black) years and the twelve (main) basins draining the western side of the Sierra Nevada to the California Central Valley in addition to the Feather River. The rivers are organized by the most northern river in the upper left to the most southern river in the lower right. The red and grey bands represent 95% confidence intervals for the regressions.

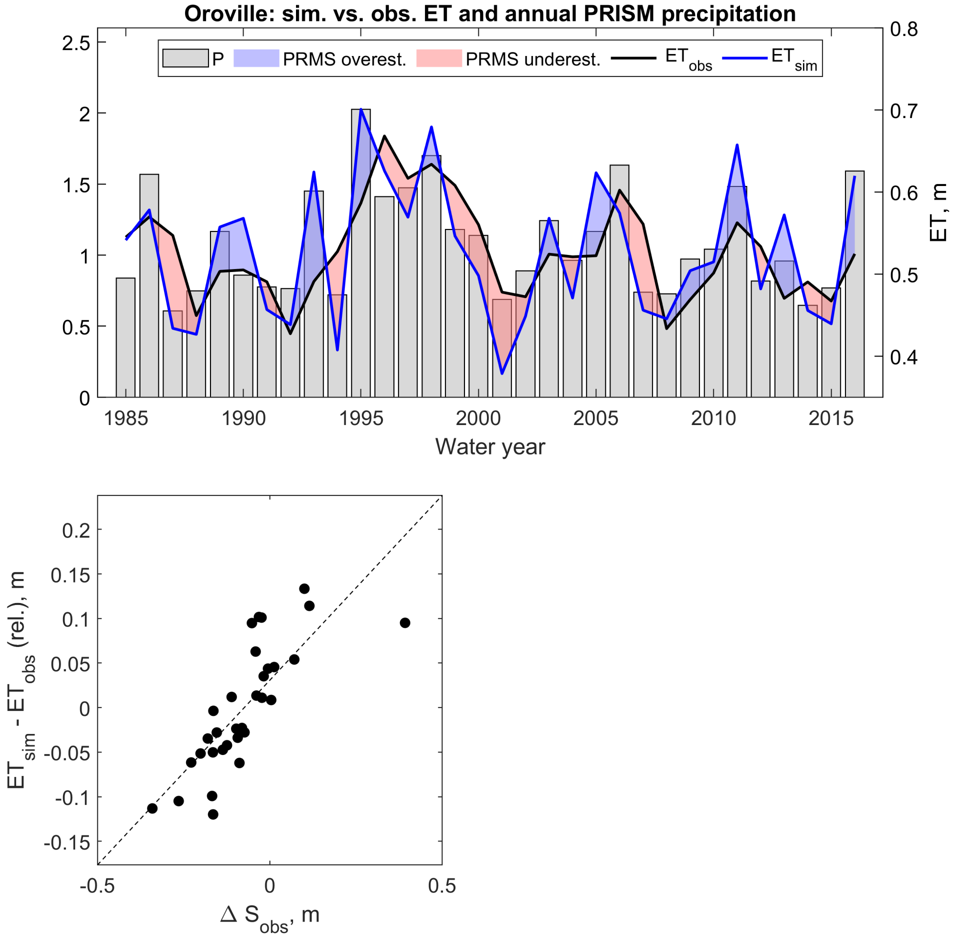
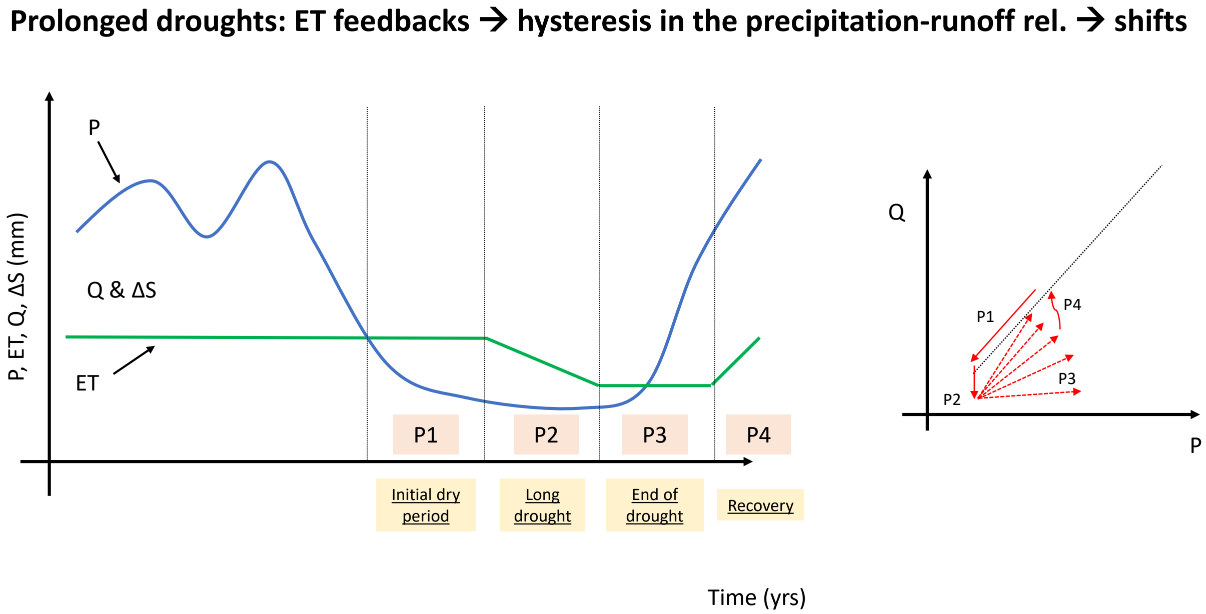


Figure 10. Top panel: simulated and ~~observed~~ estimated annual basin-wide evapotranspiration for the Feather River at Oroville (lines) and ~~observed~~ annual precipitation according to PRISM (bar chart). Bottom panel: annual differences between simulated and estimated basin-wide evapotranspiration as a function of estimated change in subsurface water storage. The absolute value of the systematic bias between simulated and observed basin-wide evapotranspiration (∼ ~~160~~ 110 mm, see Figure 7) was subtracted from observed values for readability. ~~Bottom panel: annual differences between simulated and observed basin-wide evapotranspiration.~~ Simulated annual evapotranspiration includes the groundwater-sink mass-flux component (see Section 2.3.3).



~~Precipitation-runoff relationship for drought (red) vs. non-drought (black) years and the twelve (main) basins draining the western side~~ four phases of ~~the Sierra Nevada~~ precipitation-runoff hysteresis (P1 to ~~the California Central Valley~~ P4) are discussed in ~~addition to the Feather River~~Section 4.1. ~~The red and grey bands represent 95% confidence intervals for the regressions.~~

~~Precipitation-runoff relationship for drought (red) vs. non-drought (black) years and the twelve (main) basins draining the western side~~ four phases of ~~the Sierra Nevada~~ precipitation-runoff hysteresis (P1 to ~~the California Central Valley~~ P4) are discussed in ~~addition to the Feather River~~Section 4.1. ~~The red and grey bands represent 95% confidence intervals for the regressions.~~

Figure 11. ~~Correlation across differences between simulated and observed annual evapotranspiration (relative~~ Schematic of *ET* response to ~~the systematic bias, see Figure 7)~~ climate variability and ~~four potential predictors (from~~ its effect of the ~~top~~ precipitation-runoff relationship. *ET* is initially approximated constant with time due to ~~the bottom): annual PRISM~~ its significantly smaller variability than precipitation ~~, annual maximum temperature, annual minimum temperature, and observed annual relative change in storage (also relative to the corresponding systematic bias – Figure S6)~~(**??**). ~~Regressions were calculated by separating drought and non-drought years (red and black in left column, respectively). The red and grey bands (left column) represent 95% confidence intervals for the regressions. Maximum and minimum temperature were estimated based~~ Details on ~~data in Figure 2. Simulated annual evapotranspiration includes~~ the ~~groundwater-sink mass-flux component (see Section 2.3.3).~~

~~Precipitation-runoff relationship for drought (red) vs. non-drought (black) years and the twelve (main) basins draining the western side~~ four phases of ~~the Sierra Nevada~~ precipitation-runoff hysteresis (P1 to ~~the California Central Valley~~ P4) are discussed in ~~addition to the Feather River~~Section 4.1. ~~The red and grey bands represent 95% confidence intervals for the regressions.~~